

NBER WORKING PAPER SERIES

TWO WAVES OF GLOBALISATION:
SUPERFICIAL SIMILARITIES,
FUNDAMENTAL DIFFERENCES

Richard E. Baldwin
Philippe Martin

Working Paper 6904
<http://www.nber.org/papers/w6904>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1999

We thank Assar Lindbeck, Horst Siebert, Ed Leamer, and other participants at the Kiel Week Conference for comments, corrections, and encouragement. The Swiss National Science Foundation Grant 12-50783.97 provided support for this research. The views expressed here are those of the author and do not reflect those of the National Bureau of Economic Research.

© 1999 by Richard E. Baldwin and Philippe Martin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Two Waves of Globalisation:
Superficial Similarities, Fundamental Differences
Richard E. Baldwin, and Philippe Martin
NBER Working Paper No. 6904
January 1999
JEL Nos. F1, F4, R1, F15, F22, F40

ABSTRACT

This paper looks at the two waves of globalisation (roughly 1820-1914 and 1960-present) focusing on key economic facts (trade, investment, migration, and capital flows, Industrialisation/de-industrialisation, and income convergence/divergence), economic beliefs and policymaking environments. The two waves are superficial similarities but are fundamentally different. Chief similarities include aggregate trade and capital flow ratios, and the importance of reductions in barriers to international transactions. The fundamental difference lies in the impact that these reductions had on trade in goods versus trade in ideas. Initial conditions constitute another important difference. Before the first wave, all the world was poor and agrarian. When the second wave began, it was sharply divided between rich and poor nations.

Richard E. Baldwin
Grad.Inst.of Inter'l Studies
11a, Ave de la Paix
CH-1202 Geneva
Switzerland
and NBER
Baldwin@hei.unige.ch

Philippe Martin
Grad.Inst.of Inter'l Studies
11a, Ave de la Paix
CH-1202 Geneva,
Switzerland

1 Introduction

Globalisation is creating a brave new world, according to William Greider's bestseller *One World, Ready or Not*. Page after page reveals the startling novelty of the new age, but his first chapter, "The Storm Upon Us," best encapsulates his view. "The logic of commerce and capital has overpowered the inertia of politics and launched an epoch of great social transformations," in fact, "commerce and finance have leapt inventively beyond the existing order and existing consciousness of peoples and societies."

This is glo-baloney, of course. The world has seen two waves of globalisation in the past 150 years, and in some ways, the world of 1914 was more tightly integrated than ours is today. In 1919 Keynes wrote:

What an extraordinary episode in the progress of man that age was which came to an end in August 1914! ... The inhabitant of London could order by telephone, sipping his morning tea in bed, the various products of the whole earth ... he could at the same time and by the same means adventure his wealth in the natural resources and new enterprise of any quarter of the world ... he could secure forthwith, if he wished it, cheap and comfortable means of transit to any country or climate without passport or other formality... (Keynes, 1919:6, quoted in Sachs and Warner 1995).

Globalisation it seems is not quite as new as some would have it. Indeed, many recent globalisation studies are just as unoriginal as the facts, given that they echo analysis broached in the 1960s and 1970s under the monikers of 'interdependence' and 'internationalisation'.¹

Nevertheless, the statement that there is nothing new is almost as wrong as the claim that it is unprecedented. The first globalisation wave—roughly 1870-1914—and the second—roughly 1960 to present—are superficially similar but they differ in important ways. This paper reviews some evidence on both globalisation waves, focusing on three key aspects of globalisation:

- Trade, investment, migration, and factor prices,
- Capital flows and markets, and
- Industrialisation, and income convergence/divergence.

After examining the data in the next three sections (in reverse order), we examine how economic beliefs and the policymaking environments differ between the two globalisation waves. The final section summarises the similarities and differences between the two.

The main conclusion that we take away from this exercise is that the two waves of globalisation share many superficial similarities but are fundamentally different. The chief similarities lie in aggregate trade-to-GDP and capital-flows-to-GDP ratios. These stand today approximately at the level that they attained at the end of the 19th century. Moreover, both globalisation waves were driven by radical reductions in technical and policy barriers to international transactions.² Taking a very high level of abstraction, and doing a good deal of violence to reality, we believe that one fundamental difference lies in the impact that these reductions had on trade in goods versus trade in ideas. While both waves saw reductions in both costs, the uniqueness of recent globalisation is heavily shaped by the dramatic reduction in communications

¹ See Cooper (1968, 1973), and Lindbeck (1973, 1975 and 1978); an important theme in this literature is fixed exchange rate management in the face of increased trade and capital mobility. Interestingly, Cooper in the 1960s, like Greider in the 1990s, claimed that internationalisation had outpaced the system's ability to cope. Moreover Cooper (1968:273) -- like Rodrik (1996:63, 1997:78) and Goldsmith (1994) -- recommends protection as a remedy in some cases.

² The waves were separated by the re-construction of protectionist barriers to trade as well as the imposition of capital and migration controls.

cost, what is sometimes referred to as ‘the death of distance.’ A second fundamental difference lies in the initial conditions. At the beginning of the first wave, the world was fairly homogeneous homogeneously poor and agrarian, that is. At the beginning of the second wave, the world was sharply divided between rich industrial nations and poor primary producers.

2 Industrialisation and Income Inequalities

The late 20th century globalisation wave started with a very large North-South income gap and it de-industrialised the North while industrialising the South (or some parts of it). The first globalisation wave industrialised the North, de-industrialised the South, and produced enormous income divergence between groups of nations that were not initially very far apart.

This section documents these facts, turning first to the Industrial Revolution (which drove the first globalisation wave), then to Third World de-industrialisation and finally to the income convergence/divergence issue. The section ends with a simple analytic framework that lays out one explanation of these curious contrasts.

2.1 Industrialisation and De-Industrialisation

2.1.1 The Industrial Revolution

The outcome was revolutionary but the process was evolutionary. The Industrial Revolution, which started in Great Britain, was a 100-year sequence of incremental technical, organisational, social and institutional changes. These increments, however, gradually amalgamated into a massive transformation of the British economy. The amalgamation’s starting date may be placed in the 1720s but it picked up speed as the 18th century ended and the 19th began. Placing a precise year on the Revolution is misleading; the growth acceleration was the fruit of many small transformations, not the result of a sharp policy or technological change. Be that as it may, 1776 provides as good a landmark as any since Crafts (1995) finds a structural break in the growth of British industrial production in that year. Conveniently, it is also the publication date of Adam Smith’s *The Wealth of Nations*.

The textiles and iron sectors played decisive roles in the industrial revolution, with key textiles inventions coming in the 1730-1780 period. The 1780s also saw major breakthroughs in steam engine technology. Important advances in ironmaking became widespread in the 1760s, 1770s and 1780s. Between the 1770 and the 1840, significant progress by the British machine tool industry greatly improved the precision and lowered the cost of turning iron into profitable goods such as steam engines, rails and looms.

Improved transportation was important to the Industrial Revolution. Red-letter improvements in inland water and road transportation networks appeared in the last decades of the 18th century. These expanded the internal market for manufactured goods and lowered the cost of obtaining raw materials. For instance the London-Birmingham road trip, which took two days in the 1740s, took ‘only’ 19 hours in the 1780s.

Key transport-technology advances started the first globalisation wave in approximately 1820. The most important were the rapid expansion of railroad networks (1820s-1850s), and the widespread use of steam-driven ships for inland and oceanic routes (1840s-1870s), according to Huggill (1993). Railroads revolutionised land transport by opening up vast tracts of land to the world market (earlier forms of land transport were only economic for goods with very high value-to-weight ratios). Steamship likewise revolutionised ocean travel. In the late 1830s, a top-class sailing ship from Liverpool could take up to 48 days to reach New York and 36 days to return. By the 1840s, steamships brought the normal voyage to 14 days in either direction (with much less variance). The 1870s saw further advances with the introduction of steel hulls that were lighter, stronger and required less fuel.

Finally, the whole process was fostered by the rapid development of financial intermediation (centred on in London) during the last half of the 18th century.

These advances radically re-oriented the British economy. The share of workers in industry rose progressively from 19% (1700) to 24% (1760) to 30% (1800) to 47% (1840), finally peaking at 49% in 1870 (Crafts 1989:417). The nation also shifted from a rural society to having almost two-thirds of its populations in urban areas. Although rapid total-factor-productivity growth was not remarkable, according to Nick Crafts, labour productivity did rise rapidly in some industries. From 1830 to 1860, output per hour rose 270% in cotton spinning and 708% in cotton weaving (Crafts 1989:426). During the same period,

Great Britain became a large food importer and a large exporter of industrial goods.

The growth effects of the industrial revolution look paltry next to the double-digit rates routinely observed in today's rapid industrialisers. Per capita growth rates rose from something close to zero in the first half of the 18th century to no more than 2-3% per annum in the mid-1800s. Nevertheless, even these low rates were revolutionary.

Continual growth opened the door to steady improvements in the material conditions of mankind. More importantly, industrialisation altered the fundamental character of international relations. From the beginning of recorded history until the industrial revolution, income was derived mainly from land. Wealth based on land is a zero-sum game, so violent conflicts over turf were inevitable. Wealth based on industry, by contrast, is a positive sum game—despite the fact that mercantilist and Marxist notions about competition for markets obfuscated this message for almost a century. It took two world wars to teach the lesson but the notion that more territory equals more power has been firmly relegated to intellectual history, at least in the advanced industrialised nations.

Spreading and Deepening Industrialisation

The French revolution (1789) and Napoleonic Wars (1805-1815) dominated events on the Continent during the decades bracketing the 18th century's closure. They also delayed the spread of industrialisation due to "capital destruction and losses of manpower; political instability and a widespread social anxiety; the decimation of the wealthier entrepreneurial groups; all manner of interruptions to trade; violent inflations and alterations of currency." (Landes 1969:142).

Belgium was the first to follow Great Britain into the new age, developing rapidly between 1820 and 1870. France, Switzerland, Prussia and the United States followed in the 1830s and 1840s. Industrialisation eventually spread to Russia, the Austria-Hungarian Empire, Italy, Sweden, Canada and much of the rest of Europe during the end of the 1800s. Rostow (1960) bravely, and controversially, puts dates to the various 'take-offs' (see Table 1).

As the 19th century reached into its second half, new industries and production methods emerged. Thus began the so-called second industrial revolution. In sectors such as steel, chemicals, electrical goods and engineering products based on internal combustion engines, Germany and the US leapfrogged the UK.

The price of sea and land transportation continued to fall with further advances in shipbuilding and railroads. By the 1860s, most major cities were connected by telegraph. The first transatlantic telegraph cable (1866) and the subsequent cabling of all the oceans revolutionised communications, lowering intercontinental communication times from weeks to minutes. Faster and more reliable communications spurred trade and investment. It was also critical to the development of multinationals (Dunning 1983).

Table 1: Rostow's Take-off Dates

UK	1783-1802
France	1830-60
Belgium	1833-60
US	1843-60
Germany	1850-73
Sweden	1868-90
Japan	1878-1900
Russia	1890-1914
Canada	1896-1914

Source: Rostow (1960).

2.1.2 Third World De-industrialisation in the 19th Century

While it is often forgotten, the riches of the East included much more than spices. "Before the 19th century and perhaps not much before it, some presently underdeveloped countries, notably China and parts of India, were believed by Europeans to be more highly developed than Europe" (Kuznets 1965:20). Braudel (1984) and Chaudhuri (1966) show that during the 18th century, the Indian cotton textile industry was the global leader in terms of quality, production and exports. 18th century India and China also produced the world's highest quality silk and porcelain. Before the 18th century, these manufactured goods were exported to Europe in exchange for silver since European manufactures were uncompetitive in the East (Barraclough 1978). Clearly, civilisations that invented gunpowder, paper and oceanic navigation aids were by no means primitive societies just waiting for Europe to develop.

At the end of the 19th century, however, more than 70% of Indian textile consumption is imported (mainly from Great Britain) and India instead became a net exporter of cotton (Braudel 1984). A matching, but less dramatic, story can be told for the Indian shipbuilding and iron industries. Similar cases can be found across Latin America and the Middle East (Batou 1990).

Some authors, such as Bairoch (1993), Braudel (1984) and Bairoch and Kozul-Wright (1996) claim that northern industrialisation caused southern de-industrialisation and this amplified income divergence.

“There seems little doubt that de-industrialisation in the South was the result of a massive inflow of European manufactured imports. This was particularly true of textile and clothing industries, where free trade exposed the local artisanal and craft producers to the destructive competitive gale of more capital intensive, high productivity Northern producers” (Kozul-Wright and Bairoch 1996:16).

Table 2 shows the evolution of per capita industrialisation and de-industrialisation. Note that all nations and regions started at similar levels. In 1750, all European nations were at a level between 6 and 10 (with

(UK in 1900 = 100)	<u>1750</u>	<u>1800</u>	<u>1830</u>	<u>1860</u>	<u>1880</u>	<u>1900</u>	<u>1913</u>
Developed Countries	8	8	11	16	24	35	55
Europe	8	8	11	17	23	33	45
Europe (ex-UK)	7	8	9	14	21	36	57
Austria-Hungary	7	7	8	11	15	23	32
Belgium	9	10	14	28	43	56	88
France	9	9	12	20	28	39	59
Germany	8	8	9	15	25	52	85
Italy	8	8	8	10	12	17	26
Russia	6	6	7	8	10	15	20
Spain	7	7	8	11	14	19	22
Sweden	7	8	9	15	24	41	67
Switzerland	7	10	16	26	39	67	87
UK	10	16	25	64	87	100	115
Outside Europe	7	7	11	17	33	63	116
Canada		5	6	7	10	24	46
USA	4	9	14	21	38	69	126
Japan	7	7	8	7	9	12	20
Third World	7	6	6	4	3	2	2
China	8	6	6	4	4	3	3
India-Pakistan	7	6	6	3	2	1	2
Brazil				4	4	5	7
Mexico				5	4	5	7
World	7	6	7	7	9	14	21

Source: Table 9, Bairoch (1982).

UK's level in 1900 being 100); all non-European nations were at 7 or 8, except the US, which was at 4. China and India are the chief nations represented in the Third World and they follow very similar paths dropping from something like 8 to about 3. We also see the UK's spectacular performance and the subsequent catch up of the US. Japan's industrialisation path mimics that of the world average.

Of course, the basic equality of the per capita level combined with Europe's small population meant that the Third World 'industry' dominated world production in the 18th century (Bairoch 1982 Table 10). For instance, in 1750 the Third World accounted for 73% of world manufacturing output and it continued to account for over half even as late as 1830. By 1913, however, the Third World share had dropped to a mere 7.5%.

This colonial-era de-industrialisation helps explain why many Third World countries were distrustful of unfettered international trade until very recently.

2.1.3 20th Century Industrialisation/De-Industrialisation

While the income gains of frontrunners in the 1st globalisation wave were clearly based on industrialisation, the opposite seems to hold for the 2nd wave. Indeed, apart from the handful of newly industrialised nations (NICs), convergence

Table 3: Industry's Employment Share, OECD nations, 1950-1990.

<i>% of workforce</i>	<u>1950</u>	<u>1980</u>	<u>1990</u>
Australia	36	32	26
US	36	31	28
Canada	33	33	25
Austria	35	41	37
Belgium	47	35	28
Denmark	33	31	28
Finland	28	35	31
France	35	35	29
Germany	43	45	38
Italy	29	38	32
Netherlands	40	31	26
Norway	33	29	25
Sweden	41	32	na
Switzerland	46	39	35
UK	47	38	29
Japan	23	35	34

Sources: Maddison (1989) Table C-10, World Bank (1997) Table 4.

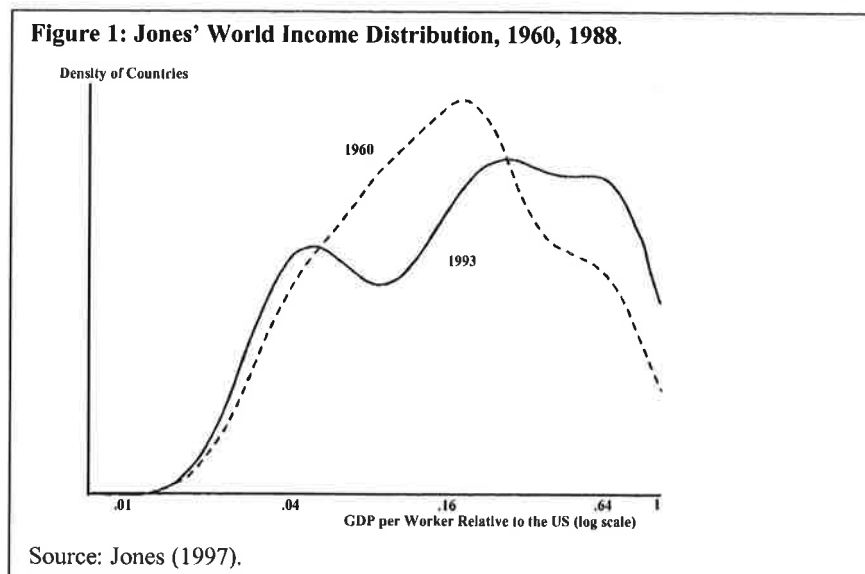
among the high-income countries has been accompanied by a marked de-industrialisation. Table 3 presents some data on this. As can be seen from the table, the share of labour force employment in industry for the OECD nations fell for most nations between 1950 and 1990. Indeed, the annual rate of de-industrialisation jumped up sharply as globalisation picked up pace in the 1980s.

2.2 Global Income Divergence in the 1st and 2nd Waves

By the time frame of human history, today's wide disparity between rich and poor countries is fairly recent. Until the industrial revolution, the whole world was poor and the logic of Malthus kept it that way. The globalisations of the 19th and 20th centuries both produced income divergence in the wider world as well as convergence among a small group of high-end nations. That is, 'twin peaks' convergence story, which has been much discussed in recent years (see Quay 1996) also occurred during the first globalisation wave.

2.2.1 20th Century Globalisation

The literature on post-war 'convergence clubs' or 'twin peaks' is highly developed and widely known, so we do no more than refer to the basic facts and provide a few references. According to Jones (1997) and



Pritchett (1997), income levels of the advanced capitalist nations converged substantially towards that of the United States while US income advanced at a pace of about 2% per annum. Additionally, incomes in a handful of NICs shot up at a truly remarkable pace, allowing them to join the ranks of the world's wealthiest nations. By contrast, incomes in many African, Latin American and some Asian nations rose slower than US incomes between the 1960s and the 1990s; others actually fell.

Figure 1 clearly illustrates this convergence and divergence story. The roughly bell-shaped 1960 distribution has become a twin-peaks distribution, with the rich nations getting richer (relative to the US) and the poor nations getting poor (again relative to the US). Jones (1997) shows that things look rosier when national incomes are weighted by population (mainly since India and China have grown faster than the US).

2.2.2 19th Century Globalisation

The paucity of data makes it more difficult to document changes in the global income distribution during the first wave of globalisation. Nevertheless, many scholars—Braudel, Kuznets, Baumol, Pritchett and Maddison, inter alios—contend that a big North-South income divergence appeared with the first Industrial Revolution.

Table 4 provides some information on the per capita income on the widening of the income distribution in Europe during the 1st globalisation wave. The first two columns show each nation's per capita income

relative to that of Great Britain's in 1860 and 1910 respectively. Slightly less than half the countries closed in on the UK. For some nations – Canada, Germany, Belgium, Denmark, France, Sweden, Switzerland and Argentina – this era marked a truly remarkable catch-up phase and the US actually surpassed its former ruler. But 11 of the 20 European nations in Bairoch's sample and all of the Asian nations in Maddison's sample lost ground. For some, such as Portugal, the drop was dramatic. The last two columns show similar calculations using data from Maddison (1995). If the table 4 data is plotted, the income distribution in 1910 is something akin to a mean-preserving spread of the 1850 distribution. The leftward shift in the lower incomes is especially marked. The Maddison data yields similar conclusions³. In short, both the 1st and 2nd waves of globalisation resulted in overall income divergence together with a top-end convergence.

The forces of history helps to shape these developments in per capita incomes, and we turn now to a simple framework that permits study of these forces of history. It is important, however, to recognise that these changes were also heavily influenced by the choices of policymakers.

2.3 An Analytic Framework

The facts on convergence/divergence and those on industrialisation/de-industrialisation are a bit puzzling.⁴ In a nutshell, we saw that the first wave of globalisation industrialised the North and de-industrialised the South. This in turn generated wide income divergence between groups that were not initially very far apart. The second wave started from a very large income gap and it de-industrialised the North while industrialising the South (or at least a good part of it). Why should globalisation first increase and then decrease income gaps? Why should this convergence U-turn be associated with de-industrialisation of the rich nations?

As it turns out, the insights of Romerian endogenous growth theory and those of Krugmanesque new economic geography provide a rigorous, if highly abstract, framework for organising our thinking about the

Table 4: Income Convergence/Divergence, 1850-1910.

(UK =100)	Bairoch data		Maddison data	
	1860	1910	1850	1913
Denmark	56	78	Canada	54 84
Germany	61	77	US	77 105
Switzerland	72	85	Argentina	56 75
Sweden	52	64	Ireland (b)	40 54
Belgium	70	80	Germany	62 76
Finland	43	47	Finland (a)	32 41
France	66	69	Sweden	55 62
Netherlands	71	72	Belgium	77 82
UK	100	100	Italy (a)	46 50
Norway	57	56	Denmark	72 75
Romania	37	35	Mexico	28 29
Russia	35	31	UK	100 100
Italy	49	44	Norway	46 45
Austro-Hungary	51	46	Austria	70 69
Bulgaria	37	31	Netherlands	80 78
Greece	41	35	France	71 69
Spain	49	41	Russia (a)	32 30
Serbia	39	29	Japan (a)	30 27
Portugal	50	33	Czechoslova	45 42
			Spain	49 45
			Pakistan (a)	22 14
			Switzerland	92 84
			China (a)	22 14
			Indonesia	28 18
			India	23 13
			Bangladesh	22 12
			Hungary (b)	54 42
			Brazil	30 17
			Thailand (b)	30 17
			Portugal	47 27
			Australia	130 109

Notes: (a) 1850 data is from 1820; (b) 1850 data is from 1870
Sources: Maddison (1995) Tables C16,D1, Bairoch (1989) Tables 1 & 4.

³ The relative decline of Russia and the populous Asian nations together with the catch-up of the US, Germany, Benelux and Nordic nations ensure a twin-peaks pattern.

⁴ Early academic contributions on the subject, such as Barro and Sala-I-Martin (1995), suggested that that income convergence was evidence for the neo-classical model whereas non-convergence was evidence for endogenous growth models. The stark-ness of this prediction has become more blurred recently, with examples of new growth models that exhibit convergence (Leung and Quah 1996) and neo-classical growth models that exhibit divergence (Baldwin 1998). This literature, however, is of little use in explaining the convergence U-turn that appears in the historical data.

economics of these facts.⁵ The framework is a stages-of-growth model with four stages and two regions (North and South) that are identical initially.

In the first, pre-globalisation stage, transport costs are high. There is little trade, and industry is primitive, rare and stagnating. Due to the high transport costs, industry is also scattered with a little bit of it in the North and in the South. This geographic dispersion itself contributes to industrial stagnation in the following way. Dispersion hinders interactions among entrepreneurs. This mutes the spillovers that could arise from the odd technological breakthrough occurring in one locale or another. The dampening of spillovers hampers innovation and technological progress and in this way, world growth is retarded.

In a second stage, when transport costs have sufficiently diminished, agglomeration forces (of the type stressed in Krugman 1991) make the even distribution of industry an unstable equilibrium. In this highly stylised world, the regions are initially identical, so the question of which region takes-off is a matter of happenstance. Whichever region edges ahead initially call it the North finds itself in a virtuous cycle. Higher income leads to a larger local market in the North and this in turn attracts relatively more investment to the North. Of course, the higher investment rate leads to a growing market-size gap and the cycle restarts. The spiral is also boosted by the localised nature of technological spillovers. Namely, Northern industry and innovators benefit proportionally more than the South from increasing industrialisation in the North. As the North experiences this stylised Industrial Revolution, Southern industry rapidly disappears in the face of competition from northern exports. In a self-generating process, the North specialises in industry and the South in primary goods.

In short, the first wave of globalisation, which is triggered by lower costs of transporting goods, generates a stark process of specialisation that both promotes and is promoted by expanding trade. The North-South income divergence, Northern industrialisation and South de-industrialisation all appear naturally.

During this first globalisation wave (stage two in the framework), the cost of exchanging goods internationally falls faster than the cost of exchanging ideas and innovations. From 1910 to the late 1960s, the rich-north-poor-south pattern was unchanged. During this stage, meant to reflect important elements of the second globalisation wave, the cost of transporting goods asymptotes towards some natural lower limit, yet the cost of 'trading' ideas continues to plummet with the cost of telecommunications. This stylised fact, which is shown clearly in Section 4.1.1, opens the door to stage four, namely the take-off of a number of developing nations.

When the cost 'transporting ideas' falls sufficiently, the core-periphery configuration becomes unstable, this time due to Krugmanesque centrifugal forces. Southern innovators and industrialist, who now have easy access to the North's technology and cheap labour, begin to close the gap. As industrial investment rises in the South, income rises, spurring local investment, and Southern industrialisation and income growth. Now it is the Southern countries turn to step on the virtuous industrialisation-income growth spiral. Northern industry suffers to some extent from the new competition. The North experiences some de-industrialisation and tends to specialise more in services. In this last stage, the two regions converge towards even income and industrialisation levels.

At first glance, this analytic framework appears to support notions of 'inequalizing trade'. In this abstract world, the big divergence between rich and poor countries is a necessary implication of Europe's Industrial Revolution and the expansion of international trade triggered both. The model, however, departs sharply from the 'inequalizing trade' paradigm one key point. While globalisation first generates massive divergence of real incomes, it subsequently becomes the driving force behind industrialisation, development and income convergence.

3 Capital Markets and Financial Integration

The popular press gives one the impression that the rapid globalisation of financial markets takes humanity into uncharted territory. This is an impression that is easy to accept given the staggering growth of international financial markets. On an average day, turnover in the foreign exchange market has grown from \$190 billion to \$1,190 billion in just ten years. Private lending flows to developing nations have increased sixfold since 1990 (IMF 1996a), and individual and institutional investors are increasingly

⁵ See Baldwin, Martin and Ottaviano (1998) for a more formal presentation.

holding foreign securities. Eichengreen (1994) points out that between 1980 and 1991, US pension funds increased the share of foreign securities in their portfolios from 1% to over 5%, and in other OECD nations the trend is similar. Finally, the world of global finance grows more intricate every year as new and highly sophisticated financial instruments appear continuously.

The popular impression of a brave new world, however, is almost entirely wrong. While there are several distinguishing features of late 20th century financial globalisation, many of its elements resemble the wave of globalisation that ended in 1914. Indeed, the only real debate among informed observers is whether we have returned to 1914-levels of financial integration (Sachs and Warner, 1995), or have yet to reach that those levels (Zevin 1989).

This section considers, in turn, the facts on capital mobility, capital market integration, and financial crisis, all of which are frequently cited as major symptoms of globalisation.

3.1 Capital Mobility

When we read that the IMF is lending record sums to Korea and Indonesia, and that private banks are topping these figures, it is easy to think of massive international lending as a new phenomenon. This is incorrect.

International financial markets in the early part of this century were highly integrated. The gold standard—combined with a dense network of financial institutions centred on London—allowed enormous flows of capital to search the world for the highest return. These flows permitted rapid growth in the ‘settler’ economies (e.g., Australia, Argentina, Canada and US) and, since much of the investment hastened the development of primary resources and vast new settlements, it fostered international trade. This, together with its stable exchange rates and unhindered convertibility, made the gold standard look like the golden age of the international monetary system.⁶

Capital mobility can be measured in different ways, and each way has its shortcomings. National current accounts are a good place to begin since a nation’s current account is by definition a measure of its net capital inflow or outflow. Table 5 (taken from Taylor 1996) provides data on the evolution of national current-account-to-GDP ratios for several countries. Since we are interested in the size of capital flows rather than their direction, the figures are the absolute value of the average current-account-GDP ratios for the various periods. Uniformly we see that numbers for the 1989-1996 period are by no means

Table 5: Capital flows since 1870.
(Average absolute value of current account as percent of GDP)

	UK	USA	Argentina	Australia	Canada	France	Germany	Italy	Japan
1870-1889	4.6	0.7	18.7	8.2	7.0	2.4	1.7	1.2	0.6
1890-1913	4.6	1.0	6.2	4.1	7.0	1.3	1.5	1.8	2.4
1919-1926	2.7	1.7	4.9	4.2	2.5	2.8	2.4	4.2	2.1
1927-1931	1.9	0.7	3.7	5.9	2.7	1.4	2.0	1.5	0.6
1932-1939	1.1	0.4	1.6	1.7	2.6	1.0	0.6	0.7	1.0
1947-1959	1.2	0.6	2.3	3.4	2.3	1.5	2.0	1.4	1.3
1960-1973	0.8	0.5	1.0	2.3	1.2	0.6	1.0	2.1	1.0
1974-1989	1.5	1.4	1.9	3.6	1.7	0.8	2.1	1.3	1.8
1989-1996	2.6	1.2	2.0	4.5	4.0	0.7	2.7	1.6	2.1

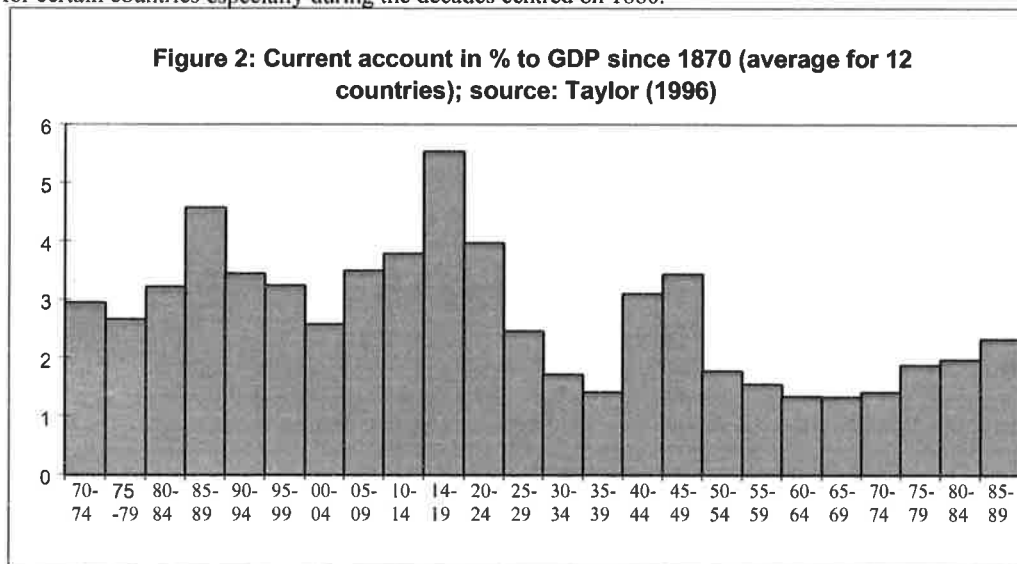
Source: Taylor (1996).

extraordinary. Indeed, except for Germany, Japan and the US, the most recent figures are noticeable smaller than the pre-WWI figures. For the UK, the average was 4.6% of GDP during the period of the Gold Standard, 1870-1914. In the 1920s, capital flows fell in most countries (to around 2.4% in the UK) and even more in the 1930s (1.1% in the UK). Capital flows reached an all time low in the 1950s and 1960s and slowly increased again starting from the 1970s to the 1990s. Although the 1990s figures are significantly higher than in any post World War II decade, they are still lower than those of a century ago.

Figure 2 illustrates the change over time aggregating data for twelve countries (those in the above table

⁶The gold standard was not perfectly stability. Southern Europe and South America nations repeatedly suspended gold convertibility (Flandreau, Le Cacheux and Zumer 1998).

plus Denmark, Norway and Sweden). Before World War I, the average size of capital flows was quite high for certain countries especially during the decades centred on 1880.



These data offer suggestive evidence that the mobility of capital has been subject to wide changes over time with a major dislocation during the Great Depression. More to the point, they suggest that the current international financial system still has a ways to go before it reaches levels of capital mobility attained during the 1st globalisation wave.

3.2 Capital and Financial Market Integration

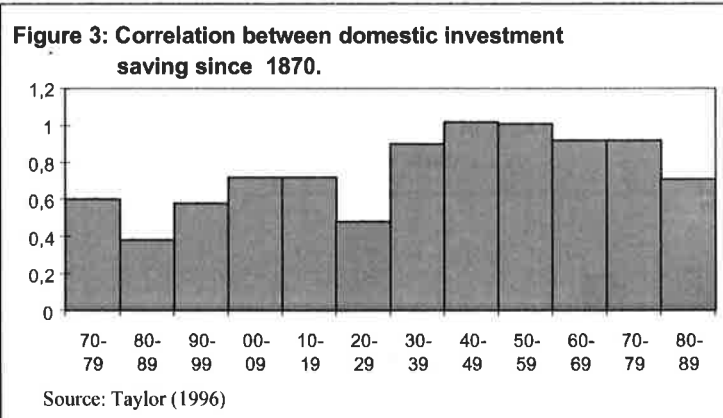
Capital flow volumes are in themselves important indicators of globalisation. Such flows, however, may give a biased image of capital market integration. On one hand, small flows are consistent with perfectly integrated capital markets as long as the returns to capital have been equalised. On the other hand, large flows are consistent with strong capital-flow restrictions if differences in rates of return are very large. To study the degree of market integration, we look at the relation between national saving and national investment over time.

3.2.1 Correlation of Investment and Savings

Feldstein and Horioka (1980) argue that in a world of perfect capital mobility, there should exist no systematic relation between domestic saving and investment since domestic saving would search the world for the highest return irrespective of the state of domestic investment.

Figure 3, taken from Taylor (1996), shows the correlation between domestic investment and saving between 1870 and 1989 for 12 major countries. The correlation displays considerable fluctuation, but since a low correlation indicates a high degree of integration, the figure presents confirms the result that the global capital market is no more integrated now than it was in beginning of the century. Indeed, the figure suggests that the late 19th century experienced significantly greater levels of integration than those observed today. More detailed evidence in Cairncross (1953:104) shows that at least for the UK, the globalisation of capital was far in excess of what we observe today; he estimates that 40% of British national saving was invested abroad in 1907. It also reveals that capital market integration has always been a long way from perfect. Econometric evidence in Zevin (1989) confirms these conclusions.

Several important historical events can be seen in the Figure 3 data. The 1880s were a period of high capital mobility characterised by intense foreign investment, some of it related to the massive expansion of railroad networks on the Continent and 'settler' countries. The crash of the early 1890s dampened flows until the boom of the 1920s⁷. Again, and more dramatic



crash of the 1930s generated a period of greater autarky. This period saw the introduction of stringent capital controls—controls that persisted into the Bretton Woods period. During the post-WWII period, these controls were gradually relaxed and eliminated in many nations. The decline of the investment-saving correlation in the 1980s is significant, but on that account, capital mobility is still lower than during the previous booms of the 1880s and 1920s.

3.2.2 Interest rates and Financial Market Integration

Evidence on nominal interest rates confirms the picture painted by figures 3. Obstfeld and Taylor (1997) look at differentials of onshore and offshore interest rates (so that the rates-of-return are not affected by pure currency risk) for short-term assets. Because the financial community is not known for systematically missing opportunities to make money without risk, this gap suggests the existence of barriers to capital mobility. In particular, Obstfeld and Taylor find that capital markets were poorly integrated in the 1930s and during the whole period of the Bretton Woods system. Differentials have become small again only in the most recent years of the flexible exchange rate regime.

Their empirical estimates also suggest a difference in the nature of financial globalisation at the end of the 19th century and at the end of the 20th century. This difference is that capital mobility at the end of the 19th century seems relatively more important for long-term than for short-term capital flows. The comparison of nominal interest differentials on US and British and short term assets between the end of the 19th century and the decade of the 1990s shows that the former were higher. This is true only for the last decade suggesting also a marked change between the 1980s and the 1990s. While this comparison is imperfect (assets in the two periods are not identical), it is consistent with other evidence. Bayoumi (1998), for instance, concludes that capital mobility is today very high for short-term instruments, and may be higher than for any other period before, but that the same cannot be said for longer-term capital movements.

Further evidence for this can be found in the gigantic foreign exchange market turnover. This turnover—several hundred trillions of dollars a year—far exceeds what would be strictly necessary to finance FDI, trade and current account imbalances. From this, and micro-studies of the foreign exchange market (e.g., Lyons, 1998), we can deduce much of the money is crossing borders many times a year. That is to say, it is invested in short-term assets. While we do not have hard data on the pattern in the late 1800s, the total size of the turnover is surely without precedent.

Bairoch and Kozul-Wright (1996) also insist on the fact that a large part of the capital flows of the late 19th century were long term in nature. Foreign direct investment (FDI) was growing rapidly during this period, accounting for as much as one-third of overseas investment. FDI was considerable in the natural resources sector but also in railways and utilities. Moreover, government borrowing was very important during this period and the flotation of new issues dominated trading of second hand debt.

This difference between the two waves can roughly be associated with the cost of trading ideas. In the

⁷ This decade is, however, marked by considerable heterogeneity across countries.

19th century, the high cost of transmitting knowledge favoured long-term capital investments. The telecommunications revolution of the late 20th century favours the rapid, almost frenetic, movement of highly liquid assets.

3.3 Financial Crises

Perhaps the most spectacular symptom of financial globalisation is the recent spat of currency and financial crises: Thailand, Korea, Indonesia, Malaysia and the Philippines in 1997-8, Mexico in 1994, and the European exchange rate crises in 1992-93. The Asian crises are especially associated with financial market globalisation. Sachs and Radelet (1998) point out: "The single most dramatic element—perhaps the defining element—of the crisis has been the rapid reversal of private capital flows into Asia."

While these financial crises sound modern and high-tech, there is very little new about them. Kindleberger (1996:203) lists no fewer than 22 crises and panics between 1870 and 1914, with the first truly international crisis coming 277 years ago. A proper characterisation of the extent and international nature of financial crises is far beyond the scope of this paper. Yet one needs only to skim a few pages of Kindleberger's book to understand how thoroughly unoriginal today's financial crises are. The 19th century crises variously involved banks, government bonds, commodities, private bonds and private equity, *inter alia*. Some affected Europe and North America, others affected developing nations. Many involved contagion-effects, pulling several nations into the panic. To give one example, the crisis of 1847-48 led to the failure of several hundred banks and companies spread across the United Kingdom, France, the Netherlands, Germany, Italy, the US and India – all within the space of 12 months.

Governmental reaction to such crises also has a long history. For instance, virtually every proposed 'solution' to today's Asian crises has a historical precedent, as Kindleberger points out in his chapters entitled "Letting it Burn Out, and Other Devices", "The Lender of Last Resort", and "The International Lender of Last Resort".

4 Trade, Investment, Migration and Factor Prices

Until the recent Asian financial crises, the rapid growth of international trade and investment was the most cited and feared symptom of globalisation. Indeed, in many circles especially in the United States—there is a firm belief that the current level of openness is unprecedented and that this takes policymakers into uncharted waters. This belief is easy to understand. US trade accounted for less than 5% of its GDP in 1960 but the current figure is more than double that. Growth in the trade-to-GDP ratio for the average OECD nation is more like 50%, but even this figure is impressive.

To informed analysts, the popular conviction that unprecedented globalisation has changed everything is wrong, or at least considerably exaggerated. As Krugman (1995:327) points out "the U.S. economy is not now, and may never be as dependent on exports as Britain was during the reign of Queen Victoria." Krugman's historical reference is correct but misses much. In terms of trade flows, the 1st and 2nd waves of globalisation are superficially similar but fundamentally different.

This section compares the two waves in terms of trade, direct investment, migration, and factor price movements.

4.1 Trade

Large-scale international trade was born the 19th century. Between the defeat of Napoleon and WWI European trade multiplied nearly forty-fold, having only doubled in the previous 100 years (Bairoch, 1989). By the end of the 19th century, a complex, but lopsided, system of international trade had taken shape. As Harley describes it:

The United States exported heavily to Europe while importing tropical raw materials, like jute, sugar and coffee, from the less developed economies of the periphery. The Continental European nations, as a group, balanced imports of temperate and tropical foodstuffs and raw materials primarily by exporting manufactured goods to Britain. Britain, in turn, earned surpluses by selling manufactured goods to the periphery, on shipping and financial services, and from large overseas investment. (Harley 1996:xii).

The UK did export manufactured goods to the US and the Continent, but these were dissimilar from the industrial goods it imported from those regions. Roughly speaking, the UK exported 'old' industrial goods

(textiles, clothing, iron, ships and railway material) while the US and Continentals focused on 'new' industrial goods (chemicals, steel and engineering goods), according to Harley (1996).

The contrast with today's trade pattern is remarkable. Developing nations are vastly less important in today's system and the commodity composition of exports across developed nations is much more symmetric. In particular, most of the world's trade (about two-thirds) takes place among rich nations that have similar factor endowments. Moreover, most of this trade (about three-fourths) is two-way trade in

Table 6: Transport Costs, 1830-1910.					Table 7: Telephone Capacity & Internet Hosts, 1986-2000.			
<i>% of Production costs</i>	<i>1830</i>	<i>1850</i>	<i>1880</i>	<i>1910</i>	<i>Thousands of voice paths</i>		<i>Thousands</i>	
					<i>Trans Atlantic</i>	<i>Trans pacific</i>	<i>Internet Hosts</i>	
Wheat	79	76	41	27.5	1986	100	41	5.1(a)
Bar Iron	92	71	33	19	1991	504	141.2	617
Manuf'd Iron Goods	27	21	10	6	1996	2'021.6	1'098.6	12'881
Cotton Thread	11	8.5	3.5	2.5	2000*	2'048.3	1'889.1	29'670a
Cotton Textile	9.5	8	4.5	2	* Projection of minimum capacities.			
Note: Figures for Hypothetical 800 km Shipment.					Source: Cairncross (1997).			
Source: Bairoch (1989:56-7).								

manufactures. Even when one zooms in to the finest level of commodity disaggregation and even when one focuses on very similar nations, one sees that world trade is dominated by intra-industry trade in similar products.

We turn now to a closer look detailed developments. We start with transport costs and trade barriers, before moving on to aggregate trade-to-GDP ratios, the geographic direction of trade and the commodity composition of trade.

4.1.1 Trade Barriers

Transport Costs in the 1st Wave

Technological innovation and infrastructure investments radically lowered late 19th century transport costs. Cheaper and faster oceanic shipping facilitated trade among coastline settlements such as London, Calcutta and New York. For instance, Harley (1980) estimates that the cost of shipping a bushel of wheat from New York to Liverpool was halved (from \$0.25) between 1830 and 1880, and then halved again between 1880 and 1914. This is only part of the story. According to his estimates for 1880, it cost as much to get wheat from Chicago to New York as it did to get it from New York to Liverpool. He shows that these inland costs also halved in both the 1830-1880 and in the 1880-1914 periods. Since wheat sold in Chicago for about \$1 a bushel throughout the 1870-1914 period, falling transport costs had an enormous impact on world wheat trade. Similar stories can be told for other bulk items such as iron and coal. Trains and ships, of course, make roundtrips, so transport improvements also fostered globalisation by expanding markets for industrial and food-producing centres.

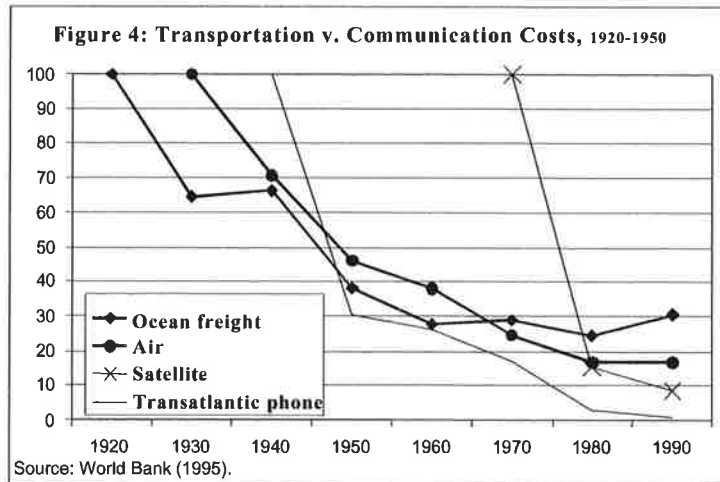
All this was led to quantitative changes for the Atlantic economies, but for nations with massive interiors—those in North American, Latin America, and central Europe – the transport improvements constituted a revolution in itself, opening up new frontiers to settlement. This, in turn, fostered the mass migrations of people and capital in the late 19th century (more on this below).

Table 6, based on Bairoch (1989) illustrates the radical impact of these innovations on trade costs. It also clearly shows how transport costs reductions were systematically more important for bulk trade commodities.

The first transatlantic telegraph cable was laid 1866 and by the turn of the century, the entire world was cabled. This was a radical change, reducing communication times from months to minutes.

Transport and Communications Costs in the 2nd Wave

In the post-war period, oceanic shipping costs continued to decline sharply up to 1960. The cost airfreight also fell dramatically, but costs flattened out in the 1980s as Figure 4 shows. The costs of communications, however, continue to plunge, as the figure shows.



Frances Cairncross further documents the remarkable decline in the cost of trading ideas in her book *The Death of Distance*.⁸ The costs of a 3-minute call from New York to London fell from about \$250 in 1930 to a few pennies today. Even as recently as 1960, the call would have cost something like \$50. The very recent expansion of telephone capacity is equally impressive, as Table 7 shows. The last column in the table also presents figures on the exponential growth of the most recent form of communications, the Internet.

Communications that are this cheap have altered the nature of international trade and investment. Cairncross (1997) is full of antidotes of the following type. Using the internet, an accountancy firm in Southern England (Dyer Partnership) acts as the finance department for a Ukrainian manufacturer of wind turbines. Dyer handles all the financial reporting, include profit and loss statements. In addition to changes in transport and communications cost, the two waves witnessed important changes in trade policy.

Although it is difficult to document, the increased ease, reliability and lower cost of telecommunications has undoubtedly promoted the explosion of foreign direct investment. This is especially true of FDI in the service sector where foreign affiliates are often selling information or expertise.

Tariffs: Then and Now

The 19th century saw both the rise of trade liberalisation and the development of modern

Table 8: Tariffs on Manufactured Goods, 1820, 1875, 1913

(Percentages)	<i>circa</i>		
	1820	1875	1913
Austria-	prohibitio	15-	13-
Belgium /a		9-10	9
Denmark	30	15-	14
France	prohibitio	12-	20-
Germany /b	na	4-6	13
Italy	na	8-10	18-
Portugal	15	20-	na
Russia	prohibitio	15-	84
Spain	prohibitio	15-	34-
Sweden	prohibitio	3-5	20-
Switzerland	10	4-6	8-9
Netherlands /a	7	3-5	4
UK	50	0	0
US	45	40-	44
Argentina	na	na	28
Brazil	na	na	50-
Colombia	na	na	40-
Mexico	na	na	40-
China	na	na	4-5
Iran	na	na	3-4
Siam	na	na	2-3
Turkey	na	na	5-10

Notes: "prohibition" indicates many import prohibitions. (a) Belgium part of Netherlands in 1820; (b) Prussian data for Germany in 1820;

Sources: Bairoch (1989) Tables 3, 5, 9; Bairoch and Kozul-Wright (1996) Tah ?

⁸ The demise of distance has been greatly exaggerated when it comes to commodity trade. The gravity model, which uses distance and partner country GDPs as its main explanatory variables, still does an excellent job of accounting for the geographical pattern of trade. In particular, the distance-trade-volume elasticity is about -0.7 ; See Baldwin (1994 Chapter 3) for further details.

protectionism.⁹ Bairoch (1989) and Harley (1996) distinguish four periods: the rise of British liberalism (1815-1846), the spread of European free trade (1846-1860), the liberal period (1860-1879), and the Continent's return to protection while Britain remained open (1879-1914). Apart from brief flirtations with free trade, non-European advanced nations, especially the US, remained protectionist until after WWII. Colonial nations that were not self-ruling had liberal policies (at least with their 'mother' countries) throughout much of this period (Bairoch 1989).

The period from 1815 and 1846 saw the establishment of Great Britain's economic supremacy and its embrace of free trade; the liberalisation of wheat imports (repeal of the Corn Laws) was the climatic event in 1846. Other European powers also moved towards freer trade during this period. This, however,

essentially entailed a shift from strict mercantilism and fragmented internal markets to modern protectionism. Germany established and freed internal trade in its *Zollverein* but raised external tariffs. Austria-Hungary, France, Russia and others generally lowered internal customs and export taxes, and shifted from prohibitions on manufactured imports to very high tariffs.

Free trade policies diffused slowly and sporadically to the Continental economies from 1846 to 1860. After 1860, liberal trade policies in

Europe spread rapidly via a system of bilateral treaties (the so-called Cobden-Chevalier treaties). Since these generally included most favoured nations (MFN) clauses, they established multilateral free trade in Europe. These liberal policies, which lasted up to the end of the 1870s, were distinctly amplified by 'natural' liberalisation, i.e. sharply lower transport cost (as discussed above).

Protectionism returned to Continental Europe after 1878 (more on this in Section 5.2). For farmers, the renewed protection just offset the sharp drop in wheat prices due to transport costs reductions (Bairoch 1989:57). For industrial goods, by contrast, the new tariffs sharply reduced or prevented increases in

Table 9: Post-Uruguay Round Protection Levels.

(Tariff equivalents by commodity,%)	<i>Med Subs S.</i>						
	<i>Nafta</i>	<i>EU</i>	<i>Jpn</i>	<i>&ME</i>	<i>Africa</i>	<i>Amer</i>	<i>Asia</i>
Grains	2	71	184	18	20	2	17
Other ag.	38	52	39	6	6	3	20
Processed food	5	13	73	6	5	2	17
Forestry	1	0	0	17	10	7	5
Fishery	1	5	3	42	8	19	11
Mining	1	0	0	17	10	4	4
Textiles	8	7	5	35	16	15	29
Clothing	19	10	9	39	20	23	21
Lumber, pulp, paper	1	0	1	24	12	9	9
Processed petro. Gds	1	1	1	17	5	12	12
Chemicals, plastics,	7	12	2	20	8	13	12
Primary steel	7	3	1	17	12	11	9
Nonferrous metals	3	1	1	24	14	7	9
Metal products	6	2	1	30	13	16	19
Transport equipment	3	5	0	25	10	19	23
Other machinery	13	7	0	24	6	19	11
Other manufactures	5	3	6	28	14	18	17

Source: Global Trade Analysis Project dataset, version 3, 1996.

Table 10: US Tariffs, 1823-1988.

	1823	1829	1842	1857	1867	1891	1908	1914	1923	1931	1
% total imports	43.4	50.8	25.3	16.3	44.3	22.9	20.1	14.9	14.1	19	
% dutiable imports	45.8	54.4	31.9	20.6	46.7	48.9	41.3	37.6	37.7	55.3	

Source: Bairoch (1993) Tab. 3.1.

industrial imports, especially from Great Britain.

Table 8 shows puts some numbers to these trends, showing that although trade barriers of the late 19th century were lower than those of the early part of the century, the 1875-1914 period was marked by high or rising barriers everywhere except the UK and the Benelux countries.

The more recent globalisation wave is also marked by tariff liberalisation. Since the signing of the GATT in the late 1940s, all developed nations have progressively liberalised tariffs and other border measures on industrial goods. Table 9 shows that except for clothing and a few sectors where special protection is important (mainly anti-dumping duties on chemicals and steel), the rich nations are quite

⁹ This section draws heavily on the masterful account of global trade policy in Bairoch (1989).

open; about as open as the most liberal European nations were in the 1870s. The table also shows that the world's remaining protection (at least the easily quantified forms of protection) is limited to two basic two categories, industrial goods in developing nations and agricultural goods in developed nations.

Table 10 provide a long-term view of the US's tariff policies. These plainly show how the United States' free trade stance is a very recent event by historical standards (apart from short-lived bursts of liberalism in the 1850s and the 1920s. The figures also clearly show the inter-war rise in protection in the US.

4.1.2 Trade Flows

Trade-to-GDP Ratios

Table 11 shows the ratios of total trade (imports plus exports) to GDP for 11 developed nations in 1870, 1910, 1950 and 1995 (actual dates for the first three columns vary slightly among nations according to data availability). The first two columns show that most nations experienced increased openness during the first wave of globalisation. The increases for Japan, Sweden and Denmark were quite spectacular, while those of other nations were more mild. The UK and Germany, for instance, had already done most of their opening up by 1860. Shares for the US, Australia and Canada actually fell during this period. A part of this stems from relative price changes between their exports (mainly primary goods, especially wheat) and domestic production but a large part is due to the fact that they never substantially liberalised their tariffs. These nations were profoundly changed by the first wave of globalisation but for them, international flows of people, capital and technology were far more important than trade (more on this below).

All countries apart from Canada saw their openness ratios drop significantly between 1910 and 1950. Moreover, all of them, except Japan, have regained or surpassed their earlier levels of trade. Thus at this level of aggregation there is very little

new about the 2nd wave of globalisation at a superficial level. However, as Lindbeck (1973) noted, the denominators in the 19th century consisted primarily of private economic activity. In modern times, however, governments account for between 30% and 50% of GDP, so the fraction of private economic activity exposed to international competition is now much greater than it was in the Victorian-era globalisation wave.

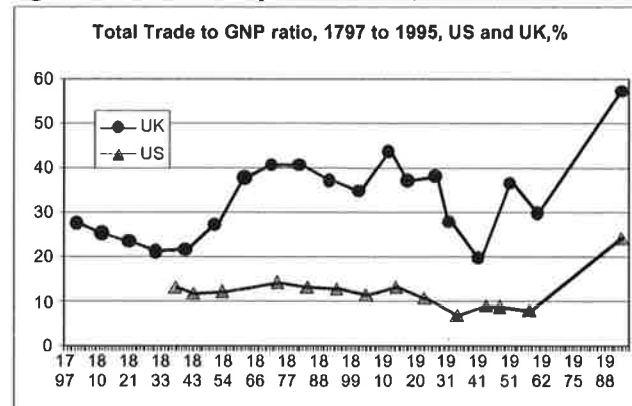
Figure 5 presents a more detailed time series for the US and the UK. The figure shows that the US is still no more open than the UK was in 1850. It also shows that the UK's burst of openness preceded the first wave of globalisation and that the US bucked the openness trend, becoming progressively more closed during the 1870-1910 period. Since WWII, however, the US has opened up remarkably. This may help explain why the 'globalisation is unprecedented' school-of-thought is so strong in the US.

Table 11: Total Trade to GDP, 1870, 1910, 1950 and 1995, Various Developed Nations. (%)

	c.1870	c.1910	c.1950	1995
UK	41	44	30	57
France	33	35	23	43
Germany	37	38	27	46
Italy	21	28	21	49
Denmark	52	69	53	64
Norway	56	69	77	71
Sweden	28	40	30	77
USA	14	11	9	24
Canada	30	30	37	71
Australia	40	39	37	40
Japan	10	30	19	17

Notes: Exact dates vary, see Kuznets (1967).

Figure 5: US and UK Openness Ratios, 1797-Present.



Source: Bairoch (1989).

Direction of Trade

As Table 12 shows, the direction of trade for Continental countries has been remarkable stable for

(% of total)		Europe	North America	South America	Asia	Africa	Other
Exports							
1860	UK	46.7	25.5	11.5	12.8	2.5	1.0
	Europe	67.5	9.1	7.7	10.0	3.2	2.5
1910	UK	35.2	11.6	12.6	24.5	7.4	8.6
	Europe	67.9	7.6	4.2	9.8	4.8	2.4
1996	UK	59.7	13.3	1.8	11.2	2.6	11.4
	Europe	76.2	7.2	2.1	10.7	2.5	1.3
Imports							
1860	UK	31.0	26.7	10.1	23.2	4.5	4.5
	Europe	61.0	14.3	7.8	12.1	3.2	1.7
1910	UK	45.1	23.8	9.1	10.3	4.8	6.9
	Europe	60.0	14.0	8.2	10.0	4.5	3.4
1996	UK	57.2	14.0	1.9	16.9	1.9	8.1
	Europe	70.7	8.5	2.9	10.5	2.7	4.7

Notes: Europe is EU in 1996.
Sources: Pre war figures. Bairoch (1974) Tables 4 and 9. 1996 figures. IMF (1997).

almost 150 years. While the deep integration of Western Europe between 1910 and 1996 has increased the regionalisation of Europe's trade, the impact is not especially large. Europe's exports to Europe rose from about two-thirds in 1860 to about three-fourths in 1996. European exports to Asia have varied little from the 10% mark and its exports to North America were unchanged between 1910 and 1996.¹⁰ There was more regionalisation on the import side between 1910 and now, but much of this has to do with the increased importance of manufactures in Europe's import basket (more on this below). d

The basic message of invariant trade shares does not carry through for the UK. As the 1st globalisation wave progressed, the UK's exports were increasingly oriented towards her colonies with India being particularly important. The share of British exports heading to the Continent fell from about half to about a third. This is partly due to rising Continental tariff barriers but also due to the rapid industrialisation of West Europe. The increased importance of Asian and African markets is especially noticeable with the former doubling to 25% and the latter increasing by half.

Composition of Trade

Manufactured goods dominate world trade today. In 1996, for instance, manufactures accounts for 73% of world trade (WTO, 1997). While low fuel and food prices boosted manufactures share significantly in recent years, the importance of manufactures has grown steadily since 1950. Importantly, the dominance of manufactures is as true for imports as it is for exports – at least in the developed nations who account for two-thirds of world trade. For instance, according to WTO (1997) manufactures make up 77.6% of European exports and 72.7% of its imports. For North America, the figures are 73.8% for exports and

	<i>Circa 1910</i>		<i>1993</i>	
	Exports	Imports	Exports	Imports
UK	75.4	24.5	97	77.2
France	59.2	25.3	78	74.6
Germany	74.5	24.4	90	73.6
Italy	38.3	38.0	89	62.9
Denmark	9.1	28.1	66	n.a.
Sweden	33.6	56.0	85	n.a.
USA	47.5	40.7	82	77.1
Canada	19.9	62.5	66	81.9
Japan	80.8	38.3	97	46.3

Sources: Kuznets (1967) Table 9 for 1910 data; World Bank (1997) and WTO (1994) for 1993 data.

¹⁰ See Anderson and Norheim (1993) for a long-term study of the regionalisation of world trade.

78.2% for imports. Japan, whose corresponding figures are 94.9% and 54.3%, is the exception but Japan accounts for only 10% of world trade.

In the late 19th century, manufactures were also important in the exports of developed nations but much less so in their imports. Table 13 presents some figures illustrating this. For the most industrialised nations, viz. UK, Germany and France, manufactures were two or three times more important in exports than imports. For the US and Italy the figures are roughly balanced but at quite a low level by modern standards. At the time, the economies of Denmark, Sweden and Canada were still heavily reliant on primary goods and this is reflected in the imbalance of manufactures in their imports and exports.

4.1.3 Trade with Emerging Economies

Some trade scholars, notably Ed Leamer, view the post-WWII globalisation as having two distinct phases. During the first phase, roughly 1950 to 1979, rapid trade expansion consisted primarily of North-North trade in manufactured goods. The developing nations played only a minor role (in terms of volume) and much of this was dominated by the exports of primary goods. During the second phase, however, trade liberalisation and rapid industrialisation changed the nature of the Third World's participation in the world trading system. In short, they became exporters of industrial goods, especially labour-intensive industrial goods. This matters, according to Leamer and Wood, since these goods compete more directly with goods produced in the developed economies. The link between this and labour market developments is considered at length below. Here we focus on the trade facts.

Imports from the emerging economies (the subset of developing nations that grew rapidly since 1970 or 1980) are still a small share of OECD GDP, about 2% for the US and about 1.5% for the OECD nations as a whole (OECD 1997). This, however, is up sharply from the ¼% it was in 1970. OECD exports to these nations have risen by approximately the same amount, but the rapid growth in imports from emerging economies has been concentrated in manufactured goods, especially labour-intensive manufactured goods. For instance, as share of total manufactured imports, the emerging economies account for 22% for US manufactured imports, up from about 10% in 1970. (The corresponding figures for the OECD as a whole are 13% and 6% respectively.)

Table 14 shows the compositional trends for developing and developed nations as a whole. Northern exports have remained primarily dominated by manufactured goods and this is as true for their exports to developing nations as it is for their exports to developed nations. The composition of the developing countries'

exports, however, has shifted rapidly from primary to secondary sector goods. While the drop in oil prices (the figures are based on current price data) exaggerates the shift, there has clearly been a significant shift in the Third World's export mix. This is especially marked after 1980; more recent data confirm this trend.

Table 14: Composition of North-South Trade, 1955 to 1989.

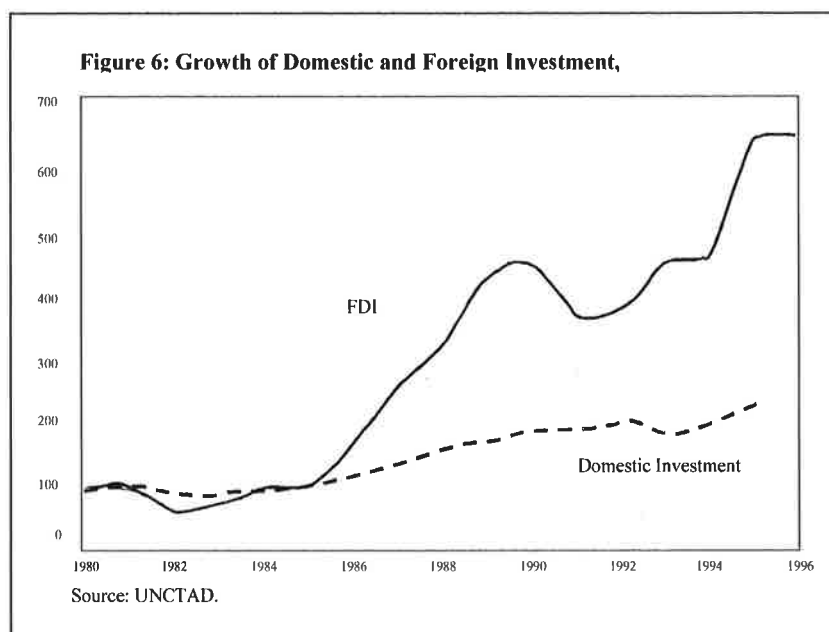
North's Exports to South:				
	1955	1970	1980	1989
Manufactures (% of total)	73	78	79	79
South's Exports to North:				
Fuels (% of total)	20	33	66	25
Other Primary (% of total)	74	51	18	21
Manufactures (% of total)	5	16	15	53

Source: Wood (1994 Table 1.1) which is based on UNCTAD data and definitions.

4.2 Multinationals and Foreign Direct Investment

Multinational corporations (MNCs) and foreign direct investment (FDI) are an essential component of late 20th century globalisation—both in terms of volume and impact. Indeed, on the microeconomic side of globalisation, FDI flows are the only thing that has changed sharply in the last twenty years. As Figure 6 shows, FDI flows rocketed in the mid-1980s. UNCTAD (1997:17), for instance, estimates that the global sales of MNCs' foreign affiliates now exceed the world trade volume by 30%. The associated footlooseness of production is also an essential component in the vague, but profound, fear of international integration that Bob Lawrence has dubbed 'globaphobia'.

Multinationals, such as the East India Company, and FDI were also an essential part of the first wave of globalisation. Indeed, in many ways FDI was more important in the first wave than in the second. Dunning



(1983) writes: “There is also little doubt that, from the viewpoint of some home and host countries, FDI, both as a channel for the transfer of resources between countries, and as a means of controlling the use of these and complementary local inputs, played a no less important role [in the 1870-1914 period] than it has done since the mid-1950s.” He guess-estimates that 35% of net capital outflows in the 1st wave consisted of FDI. The corresponding figure of the 2nd wave is harder to calculate (since the size and source of capital

Table 15: Geographical Distribution of FDI, 1914-1996, Various Home and Host Nations.

% of World Stock	1914		1960		1996	
	% of origin	% of host	% of origin	% of host	% of origin	% of host
USA	18.5	10.3	49.2	13.9	25.0	19.9
Canada	1.0	5.7	3.8	23.7	3.5	4.0
UK	45.5	1.4	16.2	9.2	11.2	10.7
Germany	10.5	6.4	1.2	6.4	9.1	5.3
France	12.2		6.1		6.5	5.2
Belgium	8.7	6.4	1.9	6.4	2.3	3.1
Italy			1.6		3.7	2.3
Netherlands	8.7	6.4	10.5	5.8	5.8	3.7
Sweden			0.6		2.4	1.3
Switzerland			3.0		5.1	1.5
Russia	2.1	7.1	na	na	0.0	0.2
Developing Nations	nil	62.8	1.0	32.3	8.9	28.4

Notes: Boxes indicate data is aggregate for the group.

Sources: Dunning (1983) Tables 5.1, 5.2 for 1914 and 1960 data; UNCTAD (1997) for 1996 data.

outflows bounced around a great deal). Nevertheless, while one might find figures like 35% for particular developing nations, any reasonable estimate of the global number would probably be less than a tenth of the 35% number.¹¹ The direction and source of FDI has also shifted significantly between the two waves.

¹¹ To reach 35% figure, a third of the average nation’s current account deficit would have to be financed by

Roughly speaking, 1870-1914 period saw FDI flowing mainly from developed countries to developing countries. In the recent wave, most FDI also originated in developed countries but has also flowed mainly to developed nations. Table 15 provides some figures on this. For instance in 1996, just four nations the US, UK, Germany, France are the source of half of the world FDI stock. Strikingly, four-tenths of the world FDI stock is also located in these same four nations. In a more complete study, Hummels and Stern (1994) show that rich nations account for 97% of direct investment outflows and 75% of the inflows.

The composition of FDI has also changed significantly. In 1914, 55% of the FDI stock was in the primary product sector, 20% in railroads, 15% in manufacturing, and only 10% in distribution, utilities, banking, etc. (Dunning 1983:89). In the late 1990s, the figures look quite different. In the EU, 63% of FDI went to services, 31% to manufactures and only 6% to the primary sector (European Commission, 1996:90). UNCTAD (1997:35) shows that only about 20% of the assets of the top 100 MNCs are in the petroleum and mining sectors.

Much of the service-related FDI is best thought of as facilitating trade in services since many types of services require a local presence. Banks are a classic example. The part of FDI that goes to the manufacturing sector, especially FDI that flows to low-wage nations, is part of another important aspect of the new globalisation, what Krugman (1995) calls the slicing up the value chain. This phenomenon was well documented in the aptly titled *The Global Factory*, a 1985 book by Kenneth Flamm and Joseph Grunwald. Because of this FDI, an enormous amount of the trade in goods such as transport equipment consists of shuttling auto parts from a company's factories in one nation to its factories in another. Krugman (1985) provides the contrasting example of a 1930s US auto plant, which "was, in effect, a facility that ingested coke and iron ore at one end and extruded passenger cars from the other."

Another striking difference is that now the global FDI pattern is quite similar to the world trade pattern. In particular, just as world trade is dominated by trade in similar goods among similar nations, FDI among rich nations is now frequently involves two-way FDI in the same industry—what has been called intra-industry FDI (Rugman 1985; Markusen 1995). Moreover, the industries in which there is much intra-industry trade among similar nations are frequently the same industries in which there is a great deal of intra-industry FDI among similar nations, according to Rugman (1985). Anecdotal evidence can be found in sectors such as transport equipment, chemicals, pharmaceutical, and processed foods.

4.3 Migration

Migration is the most obvious difference between the two waves of globalisation. Massive labour migration, often teamed with massive capital inflows, was a hallmark of the 1880-1914 period. A stylised version of these flows in the late 19th century looked as follows. The dramatic drop in transport costs – especially the construction of railroads and canals—opened frontier areas to staple production. This then generated a 20-year growth boom driven by migration, capital inflows and the export of primary goods. This pattern is sometimes called the 'Kuznets cycle'.

Table 16 shows how important migration flows were for the sending and receiving nations. The numbers are staggering by modern standards. Emigration flows equal to two to five percent of the population were entirely normal during this period.

In the 1990s, the US is the only major rich nation with high immigration rates. According to Cline (1997:85) the legal and illegal flows add to

Table 16: Decadal Migration, % Initial Population, 1880-1910, Various Nations.

% of Initial population	1880s	1890s	1900s
Senders:			
UK	-3.05	-5.20	-2.04
Italy	-1.65	-3.37	-4.87
Spain	-1.51	-6.01	-5.18
Sweden	-2.90	-7.20	-3.51
Portugal	-3.52	-4.16	-5.94
Receivers:			
US	5.69	8.94	4.02
Canada	2.27	4.89	3.71
Australia	11.28	16.59	0.77
Argentina	4.50	25.60	9.5
Brazil	1.98	3.82	8.44
N. Zealand	53.52	4.08	4.15

Notes: UK includes Ireland; dates vary slightly according data available.

Source: Green & Urquhart (1976) Table 2.

a net FDI inflow. Today, most FDI is among rich nations and fairly balanced bilaterally.

something like one million people a year, which implies a decadal migration rate of about 4% of the initial population.

4.4 Factor Prices and Globalisation

Globalisation has produced falling real wages in the US and high unemployment in Europe according to populists such as Buchanan (1998) and Goldsmith (1994) as well as from some scholars such as Leamer (1996, 1998) and Wood (1994). They point in particular to trade with low-wage nations, migration and the internationalisation of production. While these writers are at odds with the mainstream academic view (as we shall see below), they are influential. In Europe, James Goldsmith calls trade 'The Trap', and both the far right and far left have made headway by appealing to the fears of unskilled workers. In the US, fear of imports from low-wage nations was important in Congress's refusal to grant the Clinton administration trade negotiating authority in the form of 'Fast Track'. Moreover, we have seen calls for protection as a means of redressing labour market woes in virtually all industrialised nations. Protectionist appeals have been heard even from some academics such as Rodrik (1996, 1997).

The first wave of globalisation was also associated with important factor price changes, and according to some, this was responsible for anti-immigration and anti-trade legislation in the 19th century. Before looking at the 19th century facts, we briefly review recent labour market developments and evidence.

4.4.1 Factor Prices and Post-WWII Globalisation

From the end of WWII until the mid- to late 1970s, the rapid integration of industrialised nations was associated with a narrowing or stable income distribution within each OECD nation. In sharp contrast, the past two decades have been nothing less than a disaster for unskilled workers in the world's rich nations. In all OECD nations, unemployment rates are several times higher for unskilled workers, and in some nations – notably the US and UK – the earnings of these workers have fallen sharply relative to earnings of skilled workers. In the US, for one, the widening wage gap has been a zero sum game between skilled and unskilled workers since labour's total share of GDP has not budged. Consider some facts.

Some Labour Market and Income Distribution Facts

The distribution of US Family income has become much more unequal since 1980, as Table 17 (taken from Bill Cline's excellent 1997 book *Trade and Income Distribution*) shows. The US family income distribution moved little between 1970 and 1980, but became substantially more uneven from 1980 to 1992. The share of GDP received by the bottom 20% of families fell by about a sixth while the GDP share going to the highest 5% of the population rose by a sixth. The ratio of the highest to lowest incomes rose from 5.3 in 1980 to almost 10 in 1992. These sorts of figures certainly hide progress that has been made by some individuals and so they overstate the situation facing individual families. The point is that individual families can and have been working their way up through the income ladder even as more poor families are added to the lower rung. Nevertheless, the numbers are startling.

Additionally, the US wage-distribution widened during these years. As Cline (1997) documents, the ratio of the 90th to the 10th percentile wages for full time workers rose from less than 4 to

Table 17: Distribution of US Family Income, 1970, 1980 and 1992

Income Distribution Quantiles	Percentage of National Income received by each quantile			Top Income Level in Quantiles (thousands of 1992 dollars)		
	1970	1980	1992	1970	1980	1992
	0-20	5	5	4	na	18
20-40	12	12	11	na	30	30
40-60	18	18	17	na	42	44
60-80	24	24	24	na	59	64
80-100	41	42	45	na	na	na
NB:						
80-95	25	26	27	na	92	107
95-100	16	15	18	na	na	na

Source: Cline (1997 Table 1.1)

over 5 since 1970. Real wages for workers with less than High School diplomas fell by almost a quarter from 1973 to 1993. While overall US unemployment is low, for low-skilled groups the rate is at European levels, according to OECD (1997). In contrast to the US experience, the level of real wages in Europe rose (Cline 1997), but so did European unemployment rates, especially among low skilled according to OECD

(1997).

Plainly, the labour-market impact has been global. As Krugman (1997) states, rising wage inequality in the US is the flip side of the coin to Europe's rising unemployment.

Globalisation Suspects and the Evidence

The past two decades have witnessed four important changes that may help account for the labour market impact. First, trade has expanded rapidly, especially the manufactured exports of emerging economies. Second, labour-saving technology (especially factory automation and information technology) has progressed at a mighty pace, even as overall productivity growth slowed. This has reduced demand for low-skill workers in offices and factories while simultaneously boosting demand for workers who can manipulate, manage, finance and develop these sorts of technology. Third, market-oriented economic policies triumphed worldwide, weakening the power of organised labour. Finally, foreign direct investment has grown rapidly and this has fostered 'outsourcing' or 'delocation' of tasks to low wage regions.

There are by now a large number of empirical studies on the trade and wage issues (see OECD 1997 and Cline 1997 for surveys). Virtually all studies find some impact of trade on the labour market in both the US and Europe. The range of findings, however, is wide. Some find that trade accounted for virtually none of the wage gap, while other assigned 100% of the gap to trade. The consensus range is perhaps 10-20%. In the US, migration accounts for another 30-40% of drop in lowest-skilled US workers (Borjas, Freeman and Katz 1997). While these estimates cannot be simply sum up, it seem clear that maybe as much as half of the US wage gap is due to trade and migration. The rest, according to most scholars, is due to technological change.

4.4.2 Factor Prices in the First Globalisation Wave

Jeffrey Williamson (1996) and others have shown that the first globalisation wage was also associated with important factor price changes. In those days, however, the focal-point distinction was between rich landowners and poor workers rather than between skilled and unskilled workers as it is today. In turn of the century Europe, land was scarce and labour was abundant; in the New World, land was abundant and labour was scarce. Not surprisingly, labour (and capital) flowed from the Old World to the so-called settler economies in the New World. In addition to migration, sharp drops in transport costs narrowed commodity price gaps, as discussed above.

Factor rewards responded as expected (Williamson 1996). Old World wages rose relative to land values, often by a factor of 2 or 3. This was especially felt in the most open economies, such as the UK and Nordic nations. In the New World, wage-land value ratios fell by 30-50% between 1870 and 1910. Since landowners tended to be rich to start with, the ratio-shifts decreased European income inequality but probably increased inequality in the New World (Williamson 1996).

Theoretically, three factors can be seen as important determinates of the wage-land value ratios: labour migration and the attendant shift in relative factor supplies within nations, commodity price changes (via Stolper-Samuelson effects), and biased technological progress. O'Rourke, Taylor and Williamson (1996) estimate that changing labour-labour ratios and capital-deepening (due chiefly to migration and capital mobility) accounted for a quarter the changes in the New World's wage-land ratio, but for none in the Old World. Commodity price convergence was responsible for 30% of the New World's drop and about a quarter of the Old World's rise. Due to price pressures, technology tended to be labour saving in the New World and labour using in the Old. These changes accounted for about 40% of the fall in the New World ratio and about 50% of the rise in the Old World ratio.

Williamson (1996) also looks at the ratio of unskilled workers to total income per capita in an attempt to identify changes in the incomes of well-to-do individuals (capital and landowners as well as skilled workers) relative to poorer citizens (unskilled workers). He found that this proxy for domestic income inequality rose dramatically for the settler economies (Australia, Canada and US), but that it fell for labour-abundant economies in Europe (especially for pre-industrial nations such as Denmark, Italy, Norway and Sweden).

These changes in relative factor prices and incomes had an important impact on policy, according to Williamson (1996). In the US, the rise in income inequality contributed to pressures for anti-immigration legislation in the late 19th and early 20th centuries. It also fostered the anti-trade sentiments in Europe in the form of the Continent's post-1880s retreat from liberalism.

5 Economic Beliefs, Institutions and the Policymaking Environment

Keynes tells us that ideas matter more than we think, or at least used to think before he enlightened us. A full comparison of the two globalisation waves therefore requires discussion of ideas. This is especially true when attempting to distil lessons from the actions made by last century's policymakers. This section first considers differences in the macroeconomic policy environment and then turns to differences in the trade policy environment.

5.1 Macroeconomics and Finance

For late 20th century policymakers, globalisation's most startling impact concerns exchange rate management. Increasingly, officials have had to choose between floating their exchange rate (with all its attendant instability) and joining a big monetary union (with its loss of sovereignty). The adjustable peg, it seems, is dying or dead. 19th-century policymakers never faced such a stark choice. The gold standard, which was essentially an adjustable and unilaterally chosen exchange rate peg, worked quite well.¹² Our goal here is to understand how the intellectual and political climate of the late 19th century permitted this. To get there, however, requires a detour into the logic of the adjustable peg's demise.

5.1.1 Why the Adjustable Peg Died

Barry Eichengreen, who pushed the adjustable-peg-is-dead point in his 1994 book, argues that the middle ground between floating exchanges rates and monetary unions has been, or at least may soon be, completely eliminated by the speed and gigantic quantity of money that is available to challenge any fixed exchange rate. Recent events in Asia have surely made him feel more comfortable with his conclusion. Eichengreen's logic is straightforward.

A Holier 'Holy Trinity'

The 'holy trinity' of monetary policy consists of fixed exchange rates, perfect capital mobility and independent monetary policy. Governments would love to have all three, but being holy, it is unobtainable for mere mortals; the result is the well-known tri-lemma of international monetary policy. In reality, capital mobility is not perfect, so a nation with a fixed rate can get a bit of daylight between domestic and foreign interest rates. The degree of independence depends upon some combination of the restrictiveness of capital controls and the size of the country's reserves. As capital controls have eroded, the size of potential foreign exchange flows has come to eclipse all conceivable levels of official intervention.¹³ Consequently, the holy trinity has become more unobtainable. This is the first factor contributing to the adjustable peg's demise.

Self-fulfilling Currency Crises

The second contributing factor is the appearance of so-called self-fulfilling speculative attacks. These are best explained with an example.

Estonia's currency board is an adjustable peg with a peg that is very hard (politically) to adjust. Eight Estonia kroons (EEK) to the DM is the rate and Estonia promises never to change it. Estonia's inflation rate, however, is 6 or 7 times that of Germany, its monthly trade deficit is 1.5 billion EEK with an annual GDP of 63 EEK, and its interest rate is double Germany's. Now ask yourself, if no credibility were invested in the 8-to-1 rate, is this the rate Estonia would choose? If the answer is 'no' then a self-fulfilling attack is in the offing. Here are the elements.

First, speculators suspect that Estonia would devalue or float, if it were pushed off its peg. Either way, the kroon would drop, making Soros-sized profits for properly positioned speculators. Second, speculators suspect that the currency board would be abandoned if Estonian interest rates stayed at outrageous levels for long enough. Taken together, this means that boatloads of money can almost surely be made if speculators/investors are willing to pour enough money into an attack on the kroon. In a world where a trillion dollars a day changes hands, how difficult would it be to suck out the money supply of an 8 billion DM economy?

¹² The gold standard was a fixed exchange rate system where the rate could be and was changed. By definition, then, it was an adjustable peg system.

¹³ The daily foreign exchange turnover exceeds the official reserves of all IMF members combined (Eichengreen 1996).

In short, the increased holiness of the monetary holy trinity and self-fulfilling attacks increasingly force nations to choose between managed floats and membership in a big monetary union. The middle ground is rapidly vanishing.

5.1.2 The Gold Standard's Success

The gold standard was an adjustable-peg regime that worked well despite a high degree of capital mobility. The system was fully credible and there was little discussion of devaluation by the major players before 1914. In fact, currency risk was considered so minimal that foreign investment was rarely hedged. When currency fluctuations did occur, investors usually reacted in stabilising ways. And in particularly difficult crises, national central banks helped each other, as in the Baring crisis.¹⁴

How did 19th century policymakers manage this? More importantly, could 21st century policymakers do the same? At one level, the answer is simple. The gold standard worked because governments solved the tri-lemma; they were satisfied with only two elements of the holy trinity—fixed exchange rates and perfect capital mobility. Monetary policy was slaved to the exchange rate peg. At a deeper level, the puzzle is why governments at the time were happy to do this.

Eichengreen's Point

Eichengreen (1996) asserts that the gold standard “was a socially constructed institution whose viability hinged on the context in which it operated.” Specifically, the critical element was the protection that governments and central banks had from political pressure—especially pressures to sacrifice exchange-rate stability for domestic economic stabilisation. 19th century central bankers, in short, just did not face a dilemma between internal and external balance. They only cared about the exchange rate.¹⁵

Eichengreen identifies two factors as critical to this protection. First, political participation was limited (universal suffrage, for example, was unheard of) and labour unions and parties were weak. This insulated 19th century central banks from pressures to use monetary policy as an expansionary device; painful adjustments that resulted from wage contraction were always feasible since workers did not have a say.¹⁶ Second, until John Maynard published his treatises in the '20s and '30s, Keynesian economics was ill understood. The impact of monetary policy on employment, output and prices, in particular, was not fully articulated (See Box 1). Thus, ignorance of the real effects of monetary policy certainly had a positive effect on the credibility of the system.

Box 1: The Currency and Banking Schools

Having observed the impact of gold-finds on output and the inflationary consequences of debasing a currency, central bankers and economists clearly had some notion of the links between the money supply, prices, unemployment and aggregate output. Keynesianism, however, is not called a revolution for nothing. Before WWI, macroeconomics was not a distinct field and much of the debate focused on seemingly tangential issues such as whether bank credit in other forms than notes could affect prices. The pre-Keynes debate over the role money had two opposing sides, the Currency School and the Banking School. The Currency School, which included classical economists such as Ricardo, favoured strict monetary policies. They felt that supply created its own demand and that credit expansion based on ongoing business opportunities would only produce inflation. The Banking School, instead, sought credit expansion as a means of promoting economic expansion. (Based on Kindleberger, 1996:47.)

¹⁴ In this 1890 crisis, investors doubted that the Bank of England could both defend sterling and act as lender-of-last-resort for a major merchant bank. The Bank lost gold despite increases in interest rates. Assistance from the Bank of France quelled the crisis.

¹⁵ Traces of Eichengreen's point can also be found in Triffin's judgement that: “The most significant development of the [the post World War I] period was the growing importance of domestic factors as the main determinant of monetary policy.” Flandreau et al (1998) ascribes this point to Polanyi (1944).

¹⁶ See Flandreau, Le Cacheux and Zumer (1998) for a critique of this view and a detailed study of central bank independence during the gold standard.

In addition to the political insulation surrounding central bankers, the UK's dominance of the international trade and finance system was a critical factor supporting the gold standard's success. As the hegemon, Britain and her Treasury had a tremendous stake in keeping the system afloat.

Humpty-Dumpty's Fall

World War I, the rise of democracy, the end of Pax Britannica and Keynesianism all combined to shatter the social and intellectual context that supported the gold standard. The King's men and horses, so to speak, did try to put it back together again – mainly by imposing strict capital controls. They also sought to solve the internal-balance-external balance dilemma with tariffs. While the patchwork did not work, depression-era capital controls persisted in watered-down forms well into the 1980s. These helped make Bretton Woods a success since “limits on capital mobility substituted for limits on democracy as a source of insulation from market pressures” (Eichengreen, 1996, p5).

It is interesting to note that during the 1980s and 1990s, independence became a ‘must’ among central banks. One can see this, and devices such as currency boards, as a 20th century attempt to reconstruct the political insulation which 19th-century bankers enjoyed. For developing countries, the IMF may also have played this role. As some East Asian policy-makers have realised recently, it may serve their interests to have the hard choices made by a body that is insulated from local political and social pressures.

As with macroeconomics, 19th century intellectual attitudes towards trade and trade policy were quite different from those of today. We turn now to these differences.

5.2 Trade Policy

As far as trade is concerned, the first globalisation wave ended badly, with the end coming first as a whimper and then as a bang. That is, after the trade disruptions of WWI, protectionism gradually gained currency around the world, climaxing in the infamous Hawley-Smoot tariffs and corresponding retaliations.

This might be taken as suggesting that globalisation sows the seeds of its own destruction and indeed some such as Greider (1996) have claimed as much. As we shall see, however, important differences between the intellectual/policymaking environments during the two globalisation waves makes a repeat collapse unlikely. Our task here is to trace through the relevant developments in the intellectual and policymaking environments.

5.2.1 Evolution of the Intellectual Climate

The Industrial Revolution and the accompanying trade expansion emerged from an intellectual climate, called classic mercantilism, which was profoundly hostile towards imports. This belief, and the fact that tariffs and tolls were one of the few sources of public funds, meant that both internal and external trade were heavily regulated, heavily taxed or outright prohibited. Using colonies as shielded market for mother-country exports was a main ingredient in the mercantilist recipe.

In the late 1700s and early 1800s, this staunchly anti-import attitude began to change. Adam Smith, David Ricardo and others established a solid intellectual case for unilateral free trade and all the greatest economic thinkers of first globalisation came to endorse free trade. The supremacy of the free trade doctrine, however, did not hinder the development of modern protectionism, namely protectionism designed to promote development rather than simply achieve an overall trade surplus. The result was a very strong line of reasoning supporting temporary protection of manufacturing as a means of promoting the spread of England's industrialisation to the Continent and the settler economies (US, Australia, Canada, etc.). Alexander Hamilton, Friedrich List and John Stuart Mill, among others, endorsed this infant-industry protection argument. Box 2, based largely on Irwin (1996), provide a more detailed intellectual history of these developments.

Box 2: Strict Mercantilism, Laissez-Faire and Modern Protectionism

Under strict mercantilism, which dominated thinking on trade policy in the 17th and 18th centuries, imports were considered ‘bad’ and exports were ‘good’ for the simple reason that precious metals were crucial to a nation's wealth and power. Nations without access to gold

or silver mines had to obtain a balance of trade surplus in order to assure a steady inflow of gold and silver. In 1664, Mun explained: "The ordinary means therefore to increase our wealth and treasure is by foreign trade, wherein we must ever observe this rule; to sell more to strangers yearly than we consume of theirs in value" (cited in Irwin 1996:35). In short, strict mercantilism viewed trade as a means of dividing a fixed pie of world wealth. In a somewhat later form, the argument became subtler; the views being that not all imports were bad since some, such as cotton and other tropical products, were essential and unavailable at home. This subtler form of mercantilism stated that the importation of manufactures and food was 'bad', while exports, especially manufactured exports, were 'good'.

Strict mercantilist beliefs came to be challenged by the writings of a group of French philosophers, the Physiocrats, who proclaimed a harmony of private actions with public welfare (in the spirit of what Smith latter called the invisible hand). Their support for laissez-faire and free trade policies was important and influential, yet it took the genius of Adam Smith to provide the economic underpinnings of the free trade doctrine.

Adam Smith's Wealth of Nations argued that free trade maximised national wealth since it allowed a more efficient division of labour. "If a foreign country can supply us with a commodity cheaper than we ourselves can make it, better buy it of them with some part of the produce of our own industry...(Smith 1776 cited in Irwin 1996:79). Smith's treatise was the leading work on economics at the end of the 18th century and soon "all founding fathers of modern economics adopted a liberal position on international trade" according to Bairoch (1993:17). The notion of comparative costs (advantage) came to be more fully fleshed out in the first two decades of the 19th century by scholars such as Robert Torrens, David Ricardo and James Mill. The supremacy of the free trade doctrine among eminent economists, however, did not hinder the development of modern protectionism, the infant industry argument in particular.

Awareness of Britain's industrial development and rising nationalism in the early 19th century nurtured development of an intellectual argument for what we would now call infant-industry protection. Alexander Hamilton was one of the earliest and the most influential proponents of this doctrine. As a colony, US tariffs were held low to help British manufacturers 'exploit' the US market. Thus, tariff independence was an important component of US independence. Hamilton's work argued that the US should have a high tariff on manufactures and this underpinned the US' protectionist stance for more than a hundred years. Hamilton and other members of the modern protection school, concurred with Smith on the fact that protection would have static costs. Their argument was that temporary protection of industry was necessary to build up a strong economy. The ensuing growth would more than compensate the short-term loss.

On the Continent, the various writings and lectures of Friedrich List in the first half of the 1800s provided intellectual support for the protection of manufacturing, even though his most famous tract The National System of Political Economy did not appear until 1841. The intellectual respectability of infant-industry protection reached its zenith when John Stuart Mill, the pre-eminent mid-19th century economist, endorsed it in his famous 1848 book, Principles of Political Economy. Importantly, these writers typically did not recommend protection for agriculture in any nation, nor did they argue for protection in the leading industrial nation of the time, namely Britain. Furthermore, List argued that the tropical countries should embrace unilateral free trade since he thought that industrialisation was only possible in temperate climates (Bairoch 1993:18).

The writings of John Stuart Mill and Robert Torrens (who developed something akin to the optimal-tariff argument) deflowered the purity of the free trade doctrine. However, the notion that protection might improve national welfare in special cases did not provide protectionists with firm intellectual support. Several lines of argumentation buttressed the free trade doctrine (Irwin 1996:228), the strongest of which was the government-failure argument. As Edgeworth wrote in 1894, "protection might procure economic advantage in certain cases, if there was a government wise enough to discriminate those cases, and strong enough to confine itself to them...."

5.2.2 Evolution of the Policymaking Climate

As is often the case, actual policy and the beliefs of policymakers lagged the intellectual developments by a few decades. We turn now to the development of policymakers' attitudes and the pressures shaping them.¹⁷

UK's Free Trade Conversion

A glimmer of free trade policy emerged in the late 1700s. Promoted by French Physiocrats and Adam Smith's writings, the UK and France signed a liberal trade treaty in 1786 despite vociferous opposition by the French Parliament. Due to the French Revolution and subsequent wars, the Treaty had little practical effect. Moreover, the English blockade of France and France's retaliation (the so-called Continental System) greatly re-enforced policymakers' protectionist instincts on both sides of the channel.

The first sustained drive towards free trade began in the UK just after the end of the Napoleonic Wars in 1815. Starting from a position of strict agricultural protection and manufacturing tariffs of about 50%, English industry battled for free trade against the landed aristocracy. Industry embraced free trade since they believed it would boost exports by lowering British wages (industrial wages were *de facto* tied to wheat prices) and by allowing foreign nations to earn the gold necessary to pay for British manufactures.

Initially, landed interests held the upper hand, but throughout the early decades of the 1800s, UK manufacturing prowess grew both relative to its international competitors and relative to domestic agriculture. The UK's level of per capita industrialisation, which exceeded that of the rest of Europe by 110% in 1800, exceeded the Continent's average by 250% in 1830. And around 1810, the agrarian share of UK GDP exceeded that of industry by 70%, but by 1840 the roles were switched with industry's share exceeding agriculture's by 60% (Bairoch 1993:21). This shift in economic power eventually allowed the Anti-Corn Law League, formed in 1838, to push through across-the-board free trade in 1846 via the repeal of the Corn Laws.

Continental Imitation and Rejection of Free Trade

While Britain became aware of its industrial lead, Continental Europe was becoming increasingly aware of its backwardness as was seeking a means of catching up. From 1846 until 1860, Britain's continued success and the rapid expansion of trade provided a powerful show-me-the-facts justification of *laissez-faire* ideas. Starting in 1860, free trade was widely adopted by the large Continental economies via the Cobden Treaties (as discussed above). Rapid and large reductions in transport costs furthered this by lowering 'natural' trade barriers, especially for bulk items such as wheat and iron.

Soon after this liberalisation, Europe lapsed into a two-decade growth slowdown stemming mostly from a reduction in agricultural growth (Bairoch 1993:47). Per capita GDP growth, which had averaged approximately 1% to 1.3% between 1830 and 1870, slipped to 0.1% in the 1870s and 1880s (Bairoch 1993 Table 4.1). The slowdown, together with rising nationalism (the modern German, Belgian and Italian states were formed around this time), led policymakers to abandon the notion that *laissez-faire* trade policies could promote industrialisation. Across the Continent, coalitions of landed aristocracy (who suffered from falling wheat prices) and local industrialists (who suffered from British competition) pushed through protectionist barriers. Bismarck led the process. After having completed Germany's unification, he reinstated higher tariffs, using arguments provided by List: Bismarck declared that "the surfeiting of Germany with the over-production of other lands ... depresses our prices and checks the development of our industry..." (Cited in Bairoch 1989:60). The trend toward protection, which started with in Germany in 1879, reached a plateau with France's Meline tariff of 1892.

These protectionist laws, however, were not temporary and included agriculture, so they would probably not have been applauded by List, had he lived to see them. As such, this provides an early example of a sophisticated argument for protection being hijacked and misused by special-interest groups.

From Pax Britannica to Britain's Decline

Rising Continental protection, and the high levels of barriers in the settler economies, evoked a predictable response in the UK, namely calls for retaliation. This reached a peak in 1881 with the creation of a Fair Trade League that demanded protection and a retreat to trade with the Empire. Britain, however, was at the height of her hegemony; Queen Victoria ruled with a firm hand, British industrial productivity

¹⁷ This section is based mostly on Bairoch (1989, 1993) and Kindleberger (1989).

was unrivalled (UK industrial exports exceeded her industrial imports by a factor of five), and the sun never set on the Union Jack. Thus despite foreign protection, Britain chose to maintain its liberal policies and this allowed the world trading system to flourish, even as British dominance eroded (Bairoch 1989).

The First World War corresponded to a significant contraction of trade and the erection of wartime trade barriers. It also signalled the end of Pax Britannica and Britain's hegemonic support of the international trade system. The resolution of the war, the Versailles and related treaties in particular, generally ignored the health of the world trade system despite the fact that President Wilson listed global free trade as one of his famous 14 points.

Throughout the teens and twenties, protectionism gained in Europe and elsewhere in a disorganised and non-linear fashion. Exchange rate instability led to rapid changes in trade policy (trade treaties were frequently contracted for only three months) and the widespread introduction of anti-dumping and countervailing duty laws (Kindleberger 1989). Tariff rate drifted upwards despite some liberalisation of extreme wartime measures, as Table 18 shows.

Efforts were made to reverse this trend.¹⁸ For example, the 1927 World Economic Conference agreed a tariff 'truce' and subsequent reductions. The conference delegations, however, were acting in their individual capacities and were unable to get their governments to act (the notion of fast-track authority was not in currency). Consequently, almost no liberalisation occurred. The basic problem was that unlike in the 1880-1914 period Britain was unwilling and unable to unilaterally support the world trading system. "With British hegemony lost and nothing to replace it, international relations lapsed in anarchy" (Kindleberger 1989 p.167). The straw that finally broke the back of the world trading system was the 1930 US tariff hike.

The infamous Hawley-Smoot tariff finds its origin in the protectionist campaign promises made by presidential candidate Herbert Hoover to US farmers in fall 1928. A special congressional session, held in early 1929 to frame the bill, tumbled into an isolationist/protectionist spiral. The scope of tariffs was broadened to include industry, Democrats joined the Republicans and by the end "both Republicans and Democrats were ultimately pushed from the committee room as lobbyist took over the task of setting the rates" (Kindleberger 1989:170).

By 1929, the US was an economic super-power and the US market mattered enormously. The height and unprovoked nature of the US tariffs therefore attracted censure worldwide. Foreign retaliation did not wait for final passage of the bill in June 1930; Italy, France and others reacting forcefully in late 1929 and early 1930. Great Britain finally abandoned free trade, devalued sterling and instituted a system of imperial preferences a couple of years later.

By the end of the decade, the world had broken into disparate world trading systems, according to Kindleberger (1989). Germany, Italy and the Soviet Union maintained systems of bilateral trade arrangements with explicitly autarkic aims and dreams of global domination. The so-called Oslo group of small West European nations maintained regional preferences. Britain, the Dominions and Colonies were linked by the Empire preference system, and the US entertained notions of a Western Hemispheric preference system but largely stood aloof. Japan carved out her own trade system in the form of the Greater East Asia Co-Prosperty Sphere.

The breakdown of the trading system surely hastened the world towards war. It fostered acceptance of the autarkic trade philosophies of fascists in Germany, Italy and Japan. It also was used as justification for

Table 18 Average Tariff Rates on Manufactures 1913 to 1950.

	1913	1925	1931	1950
Austria	18	16	24	18
Belgium	9	15	14	11
Denmark	14	10	n.a.	3
France	20	21	30	18
Germany	13	20	21	26
Italy	18	22	46	25
Netherlands	4	6	n.a.	11
Russia	84	proh.	proh.	proh.
Spain	41	41	63	n.a.
Sweden	20	16	21	9
Switzerland	9	14	19	n.a.
UK	0	5		23
USA	44	37	48	14

Notes: "proh." indicates numerous import.
Source: Bairoch (1989).

¹⁸ Bairoch (1993) claims that the '20s saw steady or falling protection; Kindleberger (1989) rebuts this.

their territorial ambitions, with the Nazi precept of *Lebensraum*, or living space, providing the clearest example.¹⁹

This was the period when regionalism and bilateralism acquired a negative image an image that continues to this day in the minds of many trade-scholars (see, Bhagwati and Krueger 1995). Indeed, the negative experience with bilateralism in the interwar period goes a long way to explaining why multilateralism achieved such an exalted status in the post-WWII trading system.

Post WWII Institutions and Cold War Geo-Politics

By the end of the 1930s, the US and UK had learned their lessons and were attempting to liberalise trade even as WWII broke out. The war interrupted these efforts, but restoration of the world trade system showed up as one of the key objectives of the Bretton Woods institutions. Although the US Congress showed its isolationist streak by refusing to join the International Trade Organisation, it did accept the General Agreement on Tariffs and Trade (GATT). More importantly, the US has come to see the world trading system as an essential public good. That is, the world trade order is viewed as a central pillar of American prosperity, so its defence constitutes a central element of the US's self-defined national interests. Other Western nations adopted a similar attitude but war damage left them with little ability to shape events until the late 1950s.

As the Cold War heated up in the 1950s, support for the new multilateral trade system took on a geo-strategic dimension in the eyes of the US, Japan and West Europe. Expanding world trade was viewed as an important means of fortifying the 'free world.' It was also supposed to draw Third World nations into the capitalist camp, or at least to prevent them from drifting towards the communist side. Moreover, the US economic dominance put it in a unique position. Although the US import market was by far the largest in the world, the US was still an essentially closed economy until the 1970s. Thus, US tariff 'concessions' had enormous implications for her trading partners without fundamentally threatening US industry (with the usual exceptions). This was the period of Pax Americana.

Trade liberalisation in Europe was also dominated by political motives. The formation of the Common Market and the European Coal and Steel Community before it was viewed as an economic means to a political goal, that of making future European wars unthinkable.

International Trade Institutions: WTO and the EU

At the end of the first wave of globalisation, the world trade system had virtually no institutional support and indeed was not really a system in the formal sense of the word. It was supported almost exclusively by the unilateral decisions of Great Britain. Moreover, protectionism had not acquired the ill repute that it has today. 19th century policymakers from around the world (with the exception of the UK and the Low Countries) embraced protectionism as essential to rapid industrialisation. While many contemporary economists denounced this practice, the infant-industry argument received intellectual support from as eminent an economist as John Stuart Mill.

Today, we find ourselves in a very different situation. The world trade system is viewed by almost all nations as an essential public good a system that is worthy of support even for purely nationalistic reasons. With the exception of the extreme left and right, blatant protectionism is detested by policymakers and economists from around the globe and from across the political spectrum (even if the same men and women occasionally approve idiosyncratic barriers for their pet special interests).

Given the diminished dominance of the United States, it is especially important to note the strong institutions that bolstering the global trade system. Approximately one-third of world trade is conducted inside the European Union (EU). This trade is entirely free of tariffs and quotas and this status is enshrined in the Union's founding document, the Treaty of Rome, and in each member state's legislation. A supra-national legal framework, inter-national transfers and a strong political will also shore the internal market up. The EU's imports from non-European nations amount to almost another 20% of world trade. These imports are not duty free by law, but EU's tariffs are quite low on average (food imports are an important, but low volume exception). More importantly, the EU's decision-making structures in particular, the need for unanimity would make it difficult for it to raise external protection Bl a 1930s.

The GATT and WTO govern almost 70% of world trade, including the EU's external trade. A sequence

¹⁹ See the masterful account in Chapter 3 "The World through Hitler's Eyes" in Weinberg (1995).

of multilateral trade talks reduced industrial tariffs on to very low levels, and the GATT's provisions make it difficult for signatories to put these barriers back up. At the same time, WTO/GATT membership has grown from 19 nations in the late 1940s to over 120 nations today, the coverage of GATT has been expanded to include agriculture, services, and clothing. Additionally, the rules and institutions have been greatly strengthened by creation of the WTO.

Importantly, the post-WWII expansion of world trade and investment has created a powerful pro-trade lobby in every advanced industrialised nation, namely exporters and multinationals. More specifically, due to the today's very balanced pattern of trade (most trade consists of two-way manufactures trade between rich nations), the powerful export lobbies in Europe, Japan and the US have a vested interest in maintaining and advancing free trade among the three blocs. This political power, therefore, acts as a very important deterrent to another 1930s-size rise in protection rise among the advanced industrial nations.

6 Summary and the Lessons of History

A paper that covers as broad a topic as ours is difficult to summarise in a few paragraphs. At a detailed level, the differences and similarities between the two globalisation waves are covered above and we do not repeat them here. This final section focuses only on three facets of our main message, i.e., that the two waves are superficially similar but fundamentally different.

First, we would assert that trade in ideas is more important in the second wave. This can be seen in:

- The different nature of capital flows (enormous short-term flows driven by a frenetic pace of information exchange and advances in information technology rather than the long-term flows that marked the first wave).

- The different nature of FDI and MNC activity (intra-industry FDI among similar nations with a focus on manufacturing, services and outsourcing rather than the North-to-South investment in primary product sectors and railroads that were characteristic of the first wave).

- The different nature of trade (voluminous intra-industry trade among similar nations driven by scale economies and product differentiation rather than inter-industry trade driven by factor endowments differences and technology gaps as in the first wave).

- The different nature of income convergence/divergence and industrialisation/de-industrialisation. The second wave witnessed rapid (by historical standards) income convergence among leading nations coupled with their de-industrialisation and extremely rapid industrialisation of some developing nations. The first wave was marked by slow catch-up of now rich nations (to the UK) coupled with the industrialisation of converge-ers and the de-industrialisation of diverge-ers.

- The different speeds at which transportation and communication costs fell. Both dropped sharply in both waves, but the drop in communication costs far outpaced the drop in transport costs in the second wave, especially since 1980). This difference is entwined with the difference just mention, as the Section-2 analytic framework shows.

Second, the initial conditions are very dissimilar and this matters greatly. The legacy of colonialism and the vast income differences that we now see have a profound impact on the policymaking environment, especially in the Third World. On one hand, 19th century trade and de-industrialisation embittered nations, such as India and China, and this still makes it difficult for them to embrace laissez-faire trade and investment policies. On a positive side, however, the wide gap between rich and poor—teamed with more rapid technology transfer—means that those developing nations that do take-off, do so at rates that dwarf those of the 19th century.

Third, the international economic system is entirely different. Demands placed on national policymakers are far greater now than they were at the turn of the century. Electorates routinely expect expensive social welfare programs and low taxes, full-employment and low inflation, and above all, they expect steady income growth. These demands—coupled with the heightened interdependence of nations—face policymakers with challenges that were unknown in during the first wave. The tri-lemma of monetary policy is perhaps the best example of a new challenge. Another is the tendency for trade negotiations to encroach on issues that were previously considered matters of purely national concern.

The international systems also differ in a grander way. Most of the Victorian-era globalisation occurred under Pax Britannica and without any effective international institutions or formal rules. A good part of the second wave also occurred under a hegemon, this time the US, but with an important difference. The

international economic system is now bolstered by a solid set of international institutions and rules (e.g., the WTO/GATT and the IMF) and these are defended by all major nations. Such arrangements help stabilise the system, even as the US slowly shrugs off the mantle of economic leadership.

Finally, the changing beliefs of policymakers have also radically altered the international economic system. The first wave of globalisation ended badly—two global conflagrations resulted in tens of millions of deaths and incalculable material damage. The ultimate cause of this was the misguided belief that national prosperity necessarily entailed international competition for turf and/or exclusive access to markets. In the late 20th century, such notions are dismissed in advanced industrial nations. Unfortunately, these 19th century beliefs are still alive in some nations—China and Russia in particular. It would seem, therefore, that a key challenge in global governance is to draw these nations tightly into the global system of trade and investment.

7 References

- Anderson and Norheim (1993). History, geography and regional economic integration. In K. Anderson and R. Blackhurst (eds) *Regional integration and the global trading system*, Harvester-Wheatsheaf, London.
- Bairoch, P. (1974) "Geographical structure and trade balance of European foreign trade from 1800 to 1970," *Journal of European Economic History*, 3, pp 557-608.
- Bairoch, P. (1982) "International industrialisation levels from 1750 to 1980," *Journal of European Economic History*, 2, pp 268-333.
- Bairoch, P. (1989) "European trade policy, 1815-1914," in P. Mathias and S. Pollard (eds), *Cambridge Economic History of Europe, Volume VII*, Cambridge University Press, Cambridge.
- Bairoch, P. (1993). *Economics and world history*. London:Harvester-Wheatsheaf.
- Bairoch, P. and R. Kozul-Wright (1996) "Globalisation myths: Some historical reflections on integration, industrialization and growth in the world economy," UNCTAD Discussion Paper No.113.
- Baldwin, R.(1994). *Towards an Integrated Europe*. London:CEPR.
- Baldwin, R.(1998). *Agglomeration and Endogenous Capital*. *European Economic Review*, forthcoming.
- Baldwin, R., P. Martin and G. Ottaviano (1997)
- Barracough, G. (1978) *The Times atlas of world history*, Times books, London.
- Barro, R. and X. Sala-i-Martin (1995) *Economic Growth*, McGraw-Hill, NY.
- Batou, F. (1990), *Cents ans de resistance du sous-developpment*, University of Geneva.
- Bayoumi, T. (1998). Is there really a world capital market? In H. Siebert (ed.), *Globalisation and Labour*. Kiel forthcoming.
- Bhagwati, J. and A. Krueger (1995). *The Dangerous Drift to Regionalism*. Washington: American Enterprise Institute.
- Borjas, G., R. Freeman and L. Katz (1997). How much do immigration and trade affect labour market outcomes? *Brookings Papers on Economic Activity*, 1:1-67.
- Braudel, F. (1984). *Civilisation and Capitalism, 15th-18th Century: The Perspective of the World*. vol 3, New York: Harper and Row.
- Cairncross, A. (1953). *Home and foreign investment, 1870-1913*. Cambridge:Cambridge University Press.
- Cairncross, F. (1997), *The death of distance*. Orion, London.
- Chaudhuri, K.N. (1966), "India's Foreign Trade and the Cessation of the East India Company's Trading Activities", 1828-40, *Economic History Review*, 19 (2), 345-63.
- Cline, W. (1997). *Trade and Income Distribution*. Institute for International Economics, Washington.
- Cohen, Daniel, (1997), *Richesse du Monde, Pauvretés des Nations*. Paris, Flammarion.
- Cooper, R. (1968). *The economics of interdependence*. New York: McGraw-Hill.
- Crafts, N. (1989) "British industrialisation in an international context," *Journal of Interdisciplinary History*, 19, pp 415-428.
- Crafts, N. (1995), "Exogenous or Endogenous Growth? The Industrial Revolution Reconsidered," *The Journal of Economic History*, volume 55, number 4.
- Crafts, N. and C. Harley, 1992, "Output Growth and the British Industrial Revolution: a Restatement of the Crafts-Harley View", *Economic History Review*, XLV, 4, p 703-730.
- Crafts, N., 1984, "Patterns of Development in Nineteenth Century Europe", *Oxford Economic Papers*, XXXVI, 438-458.
- Crafts, N., 1989, "British Industrialization in an International Context", *Journal of Interdisciplinary History*, XIX: 3, 415-428.
- Dunning, J. (1983), "Changes in the level and structure of international production: the last one hundred years," in M. Casson (ed.) *The growth of international business*, Allen & Unwin, London.
- Eichengreen, B. (1994). *International Monetary Arrangements for the 21st Century*. Washington: Brookings.
- Eichengreen, B. (1996). *Globalizing Capital: A short history of the international monetary system*. Princeton: Princeton University Press.
- Feldstein, M. and C. Horioka (1980). Domestic savings and international capital flows. *Economic Journal* 90:314-329.
- Flandreau, M., J. Le Cacheux, and F. Zumer (1998). Stability with a pact? Lessons from the European gold standard, 1880-1914. *Economic Policy* 26: 115-162.
- Goldsmith, J. (1994). *The trap*. New York: Carroll & Graf Publishers.

- Green, A. and M. Urquhart (1976) 'Factor and commodity flows in the international economy of 1870-1914', *Journal of Economic History*, 36, pp 217-252.
- Greider, W. (1997). *One world ready or not: The manic logic of global capitalism*. New York: Simon & Schuster.
- Harley, C. (1980), "Transportation, the world wheat trade and the Kuznets cycle, 1850-1913," *Explorations in Economic History*, 17, pp 218-250.
- Harley, C. (1996), "Introduction," *The integration of the world economy, 1850-1914, Volume I*, Elgar, Cheltenham.
- Hugill, P. (1993). *World Trade Since 1431*. London: John Hopkins University Press.
- Hummels, D. and R. Stern (1994). Evolving patterns of North American merchandise trade and foreign direct investment, 1960-1990. *The World Economy*, 17:2-29.
- Irwin, D. (1996). *Against the tide: An intellectual history of free trade*. Princeton: Princeton University Press.
- Jones, C. (1997), "On the evolution of world income distribution," *Journal of Economic Perspectives*, 11, 3, pp. 3-19.
- Kindleberger, C. (1989). Commercial policy between the wars. In P. Mathias and S. Pollard (eds), *Cambridge Economic History of Europe, Volume VII*, Cambridge University Press, Cambridge.
- Kindleberger, C. (1996). *Manias, panics and crashes*. Wiley, New York.
- Krugman, P. (1995). Growing world trade: causes and consequences. *Brookings Papers on Economic Activity*, I:327-362.
- Krugman, P. (1997). *The Age of Diminished Expectations*. Cambridge: MIT Press.
- Kuznets, Simon (1965). *Economic Growth and Structure, Selected Essays*. London: Heinemann Educational Books limited.
- Landes, D. (1969) *The Unbound Prometheus*, Cambridge University Press, Cambridge.
- Leamer, E. (1998). Competition in tradables as a driving force of rising income inequality. In H. Siebert (ed.), *Globalisation and Labour*. Kiel forthcoming.
- Leung, C. and D. Quah (1996). Convergence, endogenous growth and productivity disturbances. *Journal of Monetary Economics*, 38, pp 942-547.
- Lindbeck, A. (1973). The national state in the internationalized world economy. Seminar Paper No.26, Institute for International Economics, Stockholm.
- Lindbeck, A. (1975). The changing role of the national state. *Kyklos*, 28:23-46.
- Lindbeck, A. (1978). Economic dependence and interdependence in the industrial world. In *From the Marshall plan to global interdependence*, OECD, Paris.
- List, F. (1854). *The national system of political economy*. Philadelphia: Lippincott.
- Lyons, R. (1998). *The microstructure approach to exchange rates*. Manuscript, Berkley.
- Maddison, A. (1991) *Dynamic forces in capitalist development: A long-run comparative view*, Oxford: Oxford University Press.
- Markusen, J. (1995) "The boundaries of multinational enterprises and the theory of international trade" *Journal of Economic Perspectives*, 9, 2, pp 169-189.
- Mill, J.S. (1909). *Principles of political economy*. London: Longmans.
- Obstfeld, M. and A. Taylor (1997). *The Great Depression as a watershed: International capital mobility over the long run*. NBER Working Paper No. 5960.
- OECD (1997). *Employment Outlook*. Paris: OECD.
- Polanyi, K. (1944). *The Great Transformation*. New York: Rinchard.
- Pritchett, L. (1997), "Divergence, big time," *Journal of Economic Perspectives*, 11, 3, pp. 3-19.
- Quah, D. (1996) "Empirics for economic growth and convergence," *European Economic Review*, 40.
- Rodrik, D. (1996) "Labour standards in International Trade: Do they matter and what do we do about them?," in R. Lawrence, D. Rodrik and J. Whalley *Emerging Agenda for Global Trade: High Stakes for Developing Countries*, Overseas Development Council, Washington D.C.
- Rodrik, D. (1997). *Has globalisation gone too far?* Washington: Institute for International Economics.
- Rostow, WW, (1960). *The stages of Economic Growth: A non-communist manifesto*. Cambridge University Press.
- Rugman, A. (1985) "Determinants of intra-industry direct foreign investment," in Erdilek, A. (ed) *Multinationals as mutual invaders*, Croom Helm, Kent, UK.
- Sachs, J. and A. Warner (1995), "Economic reform and the process of global integration," *Brookings Papers on Economic Activity*, I.

- Sachs, J. and S. Radelet (1998). The Onset of the East Asian Financial Crisis. HIID mimeo, Harvard University (www.hiid.harvard.edu).
- Saul, S. (1965), "The export economy 1870-1914," *Yorkshire Bulletin of Economic and Social Research*, 17, pp 5-18.
- Taylor, A. (1996). International capital mobility in history: the savings-investment relationship. NBER Working Paper No. 5743.
- Weinberg, G. (1996). *Germany, Hitler and World War II*. Cambridge: Cambridge University Press.
- Williamson, J.G. (1996) "Globalization and Inequality Then and Now: The Late 19th and Late 20th Centuries Compared", NBER Working Paper No. 5491, Cambridge.
- World Bank (1995), *World Bank Development Report*, World Bank, Washington, D.C..
- WTO (1997), *World Trade Organization, Annual Report, Volume II*.
- Zevin, R. (1992). Are world financial markets more open? In T. Banuri and J. Schor (eds), *Financial openness and national autonomy*. Oxford: Oxford University Press.

NBER WORKING PAPER SERIES

BIG-THINK REGIONALISM:
A CRITICAL SURVEY

Richard Baldwin

Working Paper 14056
<http://www.nber.org/papers/w14056>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2008

I would like to thank the follows scholars for comments and suggestions: Caroline Freund, Antoni Estevadeordal, Kati Suominen, Nuno Limao, Richard Pomfret, Alan Winters, Patrick Low and all the economists in the WTO Research Department, and the IADB's Trade and Integration Sector. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Richard Baldwin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Big-Think Regionalism: A Critical Survey
Richard Baldwin
NBER Working Paper No. 14056
June 2008
JEL No. F02,F1,F13,F15,F55

ABSTRACT

Small Think Regionalism focused on the Vinerian question: "Would a nation gain from joining a trade bloc?" Big Think Regionalism considers regionalism's systemic impact on the world trading system, focusing mainly on two questions: "Does spreading regionalism harm world welfare?" and "Does regionalism help or hinder multilateralism?" This paper synthesizes and critiques the theoretical literature in an attempt to identify the insights that are useful for thinking about regionalism's systemic impact in the new century.

Richard Baldwin
125 Nautilus Drive
Madison, WI 53705
and NBER
baldwin@graduateinstitute.ch

Big-Think Regionalism: a critical survey

Richard Baldwin

Graduate Institute, Geneva; May 2008, first draft: 28 February 2007¹

Abstract

Small Think Regionalism focused on the Vinerian question: “Would a nation gain from joining a trade bloc?” Big Think Regionalism considers regionalism’s systemic impact on the world trading system, focusing mainly on two questions: “Does spreading regionalism harm world welfare?” and “Does regionalism help or hinder multilateralism?” This paper synthesizes and critiques the theoretical literature in an attempt to identify the insights that are useful for thinking about regionalism’s systemic impact in the new century.

1. INTRODUCTION

Regionalism is sweeping the world trade system like wildfire while multilateral WTO talks proceed at a glacial pace – a correlation that leads many observers to fear that regionalism’s boon is multilateralism’s bane. This fear has pushed regionalism far up the global economic agenda and prompted a new wave of research on regionalism. But this is nothing new.

The profession’s best and brightest were focused on regionalism in the 1940s and 1950s – Jacob Viner, James Meade, Richard Lipsey, and Harry Johnson inter alia. Europe’s post-war architecture was among the world’s greatest problems and a free trade area was to be part of it, but economists were muddled over the issue.

The 1950s’ thinking straightened out the economics and established the intellectual paradigm that steered research on regionalism for decades.² The paradigm – what could be called Small-Think Regionalism – ignored systemic implications since the only large preferential arrangement – the EEC – was viewed as *sui generis*. The key question was: “Would a nation gain from joining a preferential trade arrangement?”

All this changed in the late 1980s when large-scale regionalism was re-ignited in North America. Regionalism swept the world trading system like wildfire while multilateral GATT talks proceeded at a glacial pace. In December 1990, GATT negotiations slipped into a 4-year coma causing many to fear that regionalism threatened multilateralism. As in the 1950s, regionalism was back at the top of the global economic agenda and attracting attention from the profession’s leading lights.

In 1991, Paul Krugman, Larry Summers and Jagdish Bhagwati laid out lines of analysis – what might be called Big-Think Regionalism – that continue to shape the profession’s thinking on regionalism

¹ I would like to thank the follows scholars for comments and suggestions: Caroline Freund, Antoni Esteveordal, Kati Suominen, Nuno Limao, Richard Pomfret, Alan Winters, Patrick Low and all the economists in the WTO Research Department, and the IADB’s Trade and Integration Sector. The paper was prepared for the IDB-WTO joint research project on regionalism.

² Particularly, Viner (1950), Meade (1955a,b), Lipsey (1957), Johnson (1957, 1958a, b) and Cordon (1957).

even today.³ Big-Think Regionalism focuses on the systemic implications. It comprises two principle lines of inquiry:

1. Does spreading regionalism harm world welfare?

Krugman (1991a) provided a line-sketch model that crystallised the profession's thinking around a simple comparative static exercise: "Will an exogenous variation in the number of regional trade blocs raise or lower world welfare?" He admitted, however, that this exercise missed much.

"In a fundamental sense, the issue of the desirability of free trade areas is a question of political economy rather than of economics proper. ... The real objection is ... the fear that regional deals will undermine the delicate balance of interests that supports the GATT." (Krugman 1991b, p. 23)

This led to the second line of inquiry:

2. Does regionalism help or hinder multilateralism?

Krugman (1991b, 1993) sketches a bargaining model where regionalism can help or hinder multilateralism, but it was Jagdish Bhagwati's bon mot that organised profession's thinking: do trade blocs "more readily serve as building blocks of, rather than stumbling blocks to, GATT-wide free trade." Bhagwati (1991 p. 77). Specifically, the second line of inquiry crystallised around the question of whether an exogenous variation in regional trade blocs made multilateral tariff cooperation more or less likely.

The two main lines of inquiry embraced the notion that one could reasonably view changes in the number of blocs as exogenous. A third line of inquiry extended the issue by endogenising regionalism's spread. The focus was on cause and extent of spreading regionalism; it turned on positive political economy question of which trade blocs would actually emerge (Baldwin 1993, 1995, Grossman and Helpman 1995, Yi 1996, Freund 2000a, b, Aghion, Antras and Helpman 2004).

Summers (1991) contribution was to stake out one extreme of the debate on both lines of inquiry. He argued that "plausible" regional arrangements were natural trade blocs and thus would raise world welfare. He also asserted that reasonable regional arrangements were as likely to accelerate the general liberalization process as to slow it down. Hence his famous assertion that all the 'isms are good: unilateralism, bilateralism, plurilateralism and multilateralism. Summers rejected the notation that regionalism and multilateralism were enemies as Krugman, Bhagwati and many suspected; regionalism and multilateralism were the two legs on which the world was walking towards global free trade. In retrospect, Summers was closer to the mark since the Uruguay Round finished in 1994, securing enormous advances in the breath and depth of multilateralism despite, or even because of, spreading regionalism.

Plan of the paper. The goal of this paper is to summarise and evaluate the Big-Think regionalism literature. My focus is almost exclusively on theoretical work due to length limits. Before turning to the literature, Section 0 presents the elemental economic effects that concern Big-Think Regionalism in a way that helps to fix ideas and terms. Section 3 considers the stumbling/building block issue and Section 4 looks at the "Is bilateralism bad" literature. The third line of inquiry is covered in Section 5. A summary and concluding remarks are in Section 6.

³ See particularly, Krugman (1991a,b), Bhagwati (1991 Chapter 5), Summers (1991), but also Krugman (1993) and Bhagwati (1993).

2. BASIC ECONOMIC EFFECTS

Before launching into a review of Big-Think Regionalism, it proves useful to conduct a quick review the basic economics of preferential trade liberalization as far as the political economy interaction between regionalism and multilateralism is concerned. There is nothing new in this review and it ignores many elements at that important economically – e.g. scale effects, growth effects and location effects. It is necessary since the literature is marked by a conceptual ‘spaghetti bowl’ – a tangle of conflicting, overlapping and competing terminologies for basic effects. My sole aim here is to establish a common set of labels and notation for the key concepts underpinning thinking in the Big-Think Regionalism literature.

There are only three core effects. All have been known at least since 1950. To avoid creating yet another set of terms, I label them according to their intellectual fathers.

2.1. Smith’s certitude, Haberler’s spillover and Viner’s ambiguity

The theory of preferential trade did not begin in 1950 with Jacob Viner.⁴ Early contributions include Smith (1776), Taussig (1892), and Torrens (1844). Adam Smith’s contribution highlights one of the most robust finding in the field – what might be called ‘Smith’s certitude’:

When a nation “exempt[s] the good of one country from duties to which it subjects those of all other ... the merchants and manufacturers of the country whose commerce is so favoured must necessary derive great advantage.” (Smith, 1776 as quoted in Pomfret 1997).

Much later Gottfried Haberler asserted that all members of a preferential trade agreement (RTA) must gain while third nations must lose.⁵ The first part of the assertion is wrong, but what might be called ‘Haberler’s spillover’ – the part about third nations losing – turns out to be almost as robust as Smith’s certitude. Haberler’s spillover and Smith’s certitude are the linchpin’s of the Big-Think regionalism discussion.

The only basic element added in the post-war period came with Jacob Viner’s famous 1950 book *The Customs Union Issue*. Viner’s key finding is that discriminatory tariff liberalization has ambiguous welfare effects (‘Viner’s ambiguity’).⁶ Viner’s ambiguity is quite general but one is hard pressed to see this from the analysis in his book. An RTA is nothing more than a special case of non-uniform commodity taxation, but Viner did not have the benefit of modern economic tools for tax analysis.

⁴ See the excellent survey, Pomfret 1997 Chapter 8.1, on pre-Vinerian contributions, also Pomfret (1986) and O’Brien (1976),

⁵ His discussion runs over several pages but here he asserts it fairly directly: “There is no difference in kind, but only one of degree, between the grant of lower preferential duties upon imports from certain country and a general reduction in tariffs. A partial reduction is better than none at all (although, of course, a general reduction would be still better, from an economic standpoint).” (Haberler 1936 p. 384).

Haberler’s spillover was certainly understood by scholars before Haberler (e.g. Bismarck used this aspect of customs union to force/cajole many German-speaking states to join his unified Germany), but I assign it to Haberler since Haberler’s 1936 book shows that mainstream trade economists were confused about the theory of the second best. This illustrates why Viner’s 1950 book was viewed as such a landmark.

⁶ Viner’s consideration of other effects of customs union formation (its impact on the terms of trade, economies of scale, cartels, administrative efficiency, the pressure to harmonize excise taxes, and the necessity to go beyond tariff removals in order to remove trade barriers) made much less lasting impact on the literature, but remains a fascinating and highly accessible read.

Rather, he conducted the analysis using the enduring but imprecise concepts of 'trade diversion' and 'trade creation'.

"The analysis will be directed towards finding answers to the following questions: in so far as the establishment of the customs union results in change in the national locus of production of goods purchased, is the net change one of diversion of purchase to lower or higher money-cost sources of supply, abstracting from duty-elements in money costs ... If the customs union is a movement in the direction of free trade, it must be predominately a movement in the direction of goods being supplied from lower money-cost sources than before. If the customs union has the effect of diverting purchases to higher money-cost sources, it is then a device for making tariff protection more effective. None of these questions can be answered a priori, and the correct answers will depend on just how the customs union operates in practice." (Viner 1950 p.44).⁷

'Trade diversion' and 'trade creation' are misleading since they suggest trade volumes are the key even though his words clearly indicate that cost/price changes are what matter. Moreover they fail to cover all the effects generated by discriminatory tariff liberalisation – even in a simple Walrasian setting. Given these shortcomings and the decades-long debate on ‘what did Viner really mean’ (a debate in which Viner himself participated without notable effect), it is curious that the terms have enjoyed such enduring success.⁸

The generality of Viner’s ambiguity is glaringly obvious to readers schooled in the theory of the second best (preferential liberalisation induces new distortions while removing others) but Viner’s book was a landmark. The theory-of-the-second-best was unknown in 1950 and many of Viner’s contemporaries – Haberler, for example – were muddled over the key differences between general and preferential liberalisation.

2.1.1. Kemp-Wan logic

A fourth elemental effect in the regionalism literature concerns the interaction between preferential and multilateral tariff cutting. It is not really a basic economic effect but rather a specific combination of effects motivated the fact that the most important regional liberalisation over the last 60 years (Europe and North America) has been accompanied by multilateral liberalisation. When thinking about this teaming of multilateral and regionalism liberalisation, the guiding light is the Kemp-Wan logic.

Meade (1955) introduced analysis that produced one of the few general statements that can be made about RTAs – the Kemp-Wan theorem.⁹ Kemp-Wan (1976) demonstrate that RTAs could be designed to be Pareto improving for every member of the RTA and the world at large. The logic is elegant. Assume two nations sign a RTA and alter their external tariffs to freeze their external trade flows; the external trade flows can then be treated as part of the bloc’s endowment. Removal of all intra-RTA

⁷ Viner (1950) is worth reading in the original. His but informal reasoning is full of insights and it anticipates much of the economic and political economy theory as well as the political economy debates that have surrounded economic integration in the subsequent six decades. The key passages are reproduced verbatim in Box 2.

⁸ The basic problem was the profession found the simple trade creation/diversion paradigm to be effective in communicating the crucial welfare-ambiguity result but the words did not fully capture all the basic economic effects. Arvind Panagariya, for example, suggests that the terms persist since they are “highly effective tools of focusing policy makers’ attention on the ambiguous welfare effects of RTAs.” (Panagariya 1999).

⁹ Meade identified the basic result in 1955 when he argued that were it not for external trade considerations (the PTA's trade with non-members), duty-free internal trade would be optimal: "if all trade barriers take the form of fixed and unchanged quantitative restrictions, then a customs union must increase economic welfare" (Meade 1955b p. 98). As often happened in ‘customs union theory’, the result was re-invented repeatedly (Vaneck 1965, Ohyama 1972, and Kemp and Wan 1976). The profession knows it as the Kemp-Wan theorem. Recently, Krishna and Panagariya (2000) follow-up Meade’s insight that the key is to freeze external tariff vectors to show that Kemp-Wan holds for FTAs as well as customs unions, if the FTA is free to choose all external tariffs.

barriers thus shifts the two-nation bloc from a second-best situation to a first-best situation (i.e. laissez-faire in goods and factors given tastes, technology and endowments). The first welfare theorem of Walrasian economics guarantees an increase in economic efficiency and lump-sum transfers within the RTA ensure welfare gains for all. Third nations' are unaffected since their trade vectors are unaffected. Dixit and Norman (1980) generalise the analysis, showing that the Kemp-Wan improvement can be obtained without lump sum transfers; intra-RTA commodity taxes and subsidies are sufficient.

Of course, real-world RTAs do not adjust external tariffs in a Kemp-Wan manner, nor do they have access to large lump-sum transfers. Nevertheless, the theorem is important from a policy perspective. It proves that RTAs are not necessarily bad for world welfare. Moreover, it helps us think about why the duo of multilateral and preferential tariff cutting – a duo that has been in operation since the 1950s – may have been critical to explaining why post-WWII regionalism has had a relatively benign impact on the world trade system to date. Certainly much more benign than the European regionalism between the wars.

2.2. Modern treatment of Viner's ambiguity

The first economists to apply modern economic analysis to Viner's question were Meade (1955a,b), Lipsey (1957) and Gehrels (1956-7). Modern tax analysis shows that the welfare impact of any tax change is captured by two terms in the Walrasian setting – one related to the change in consumption over the tax wedge, the other related to the level of consumption times the change in the actual price paid. James Meade's pioneering analysis applied this to import taxes (tariffs) where the two terms may be called the trade volume and trade price effects (the trade price effect is often called the terms-of-trade effect). For the nation imposing the tariff, the net welfare effects is related to the initial tariff wedges on bilateral trade, changes in bilateral imports, and the changes in bilateral border prices according to:

$$(1) \quad \text{Net Home welfare effect} = (p-p^*)dM - (M)dp^*$$

where p and p^* are the vectors of internal and border prices, M is the vector of bilateral imports (exports are negative imports), and dM and dp^* are the vector of changes in bilateral trade volumes and border prices, respectively. See Box 3 a derivation of this expression in a simple linear case, and the appendix for a more general demonstration that allows for a variety of other effects (pro-competitive effects, scale effects, location and accumulation effects) in a more general economic setting.

Viner's special case: adding up created and diverted trade. An antiquated but enduring rule-of-thumb for evaluating RTAs turns on the volume of trade created and diverted. We can use Meadean analysis to show exactly what is being assumed away. Assuming tariffs are the only barriers (no export taxes or subsidies so $p-p^*$ equals the vector of bilateral tariffs, T) and ignoring changes in all border prices (so Mdp^* is zero), the welfare effect boils down to T times dM . Further assuming that tariffs are identical on all imports, the net welfare effect is proportional to the sum of changes in imports. In words this says a RTA member gains if the RTA creates more trade than it diverts. A slightly more general test – one that allows for different bilateral tariffs – is to check the change in tariff revenue collected. Of course this test ignores Smith's certitude and Haberler's spillover, but it was the best economists could do in the 1950s and 1960s without computers.

Meade's primary, secondary and tertiary effects. Meade (1955b) described the trade volume and border price effects as the primary effects, but he listed two other categories of effects. Meade's secondary effects are the substitution and income effects of a tariff change on other markets. His tertiary effect concerns general equilibrium adjustments necessary to insure the balance of payment.

2.3. Illustration of basic economic effects

Smith's certitude, Haberler's spillover, Viner's ambiguity capture most of the basic economics of RTAs, and, together with and the Kemp-Wan logic, most of the political economy reasoning in the big-think regionalism literature.¹⁰ It is possible to deal with these mathematically, however to demonstrate the basic interactions among the elemental effects, and to facilitate the exposition of the logic of the big-think regionalism literature in the sequel, it is useful to have an amenable and flexible analytic framework, especially one that lends itself to graphical analysis.

The simplest framework that meets these requirements is a Walrasian 3-nation model (Home, Partner and RoW) with 3-goods (goods 1, 2 and 3); each nation exports two goods and imports the other good (Figure 1). Since each nation has two sources of imports, tariff discrimination can be a real issue in all markets. To rule out Meade's secondary effects, tastes are assumed to be identical across nations and additively separable in all goods.¹¹ For simplicity's sake, the three nations are symmetric in terms of size and the MFN tariff they initially impose.¹²

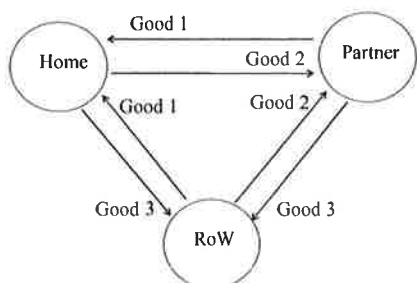


Figure 1: The RTA diagram's trade pattern

The two trading equilibrium (regionalism versus multilateral free trade) in a typical market (good 1) can be worked out with the help of the 'RTA Diagram' (Figure 2). The analyses for imports of goods 2 (into Partner) and 3 (into RoW) are isomorphic due to the strong symmetry.

The diagram shows the export supply curves (marked XS with the appropriate superscript indicate the origin nation) for Home's two potential suppliers (two leftmost panels). The horizontal sum of the XS curves is shown in the rightmost panel (as MS_{FT}) along with Home's import demand curve, MD. Under global free trade, the domestic and border price is P^{FT} as shown in all nations for all goods. Assuming all nations impose a specific tariff T on an MFN basis, the internal price in Home is driven up to P while the border price is driven down to $P-T$ for both suppliers. Home imports drops with the reduction divided equally among the two suppliers.

From MFN tariffs to FTA. If an FTA or customs union is formed between Home and Partner, the total import supply curve becomes the kinked MS_{FTA} curve. The resulting internal price fall to P' but there are now two border prices. The FTA raises the price facing Partner exporters from $P-T$ to P' while it lowers the RoW border price from $P-T$ to $P'-T$. Partner exports expand while RoW exports contract. Identical things happen in the market for good 2, but here Home is the exporter and Partner the importer. Nothing happens in the market for good 3 (where RoW is the importer) since RoW

¹⁰ Of course when considering the full economic impact, one must consider scale economies, pro-competitive effects, variety effects, location effects and growth effects. (Baldwin and Venables 1995). Most of these, however, are not critical in the 'big think' regionalism literature.

¹¹ The reader can mentally insert a 4th untaxed good that enters the utility function linearly to formally eliminate Meade's tertiary effects.

¹² The RTA diagram can be thought of as a modification of the Blackhurst (1972) diagram.

maintains its MFN tariff and the strong separability assumptions rule out Meade's secondary and tertiary effects.

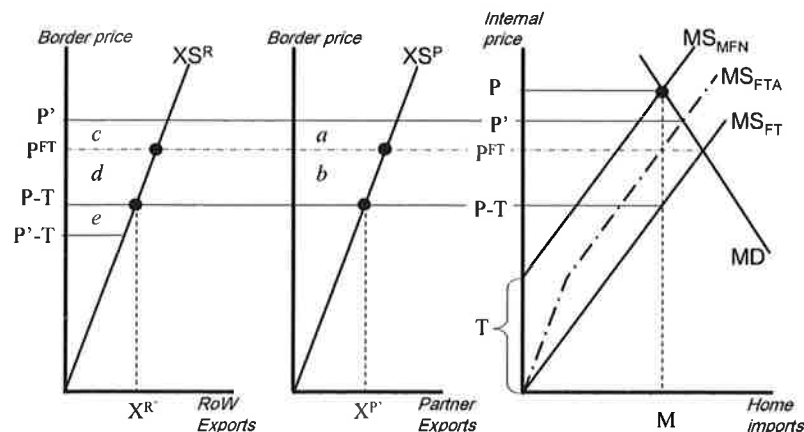


Figure 2: The RTA (Preferential Trade Arrangement) Diagram

Source: Baldwin and Wyposz (2003 chapter 5).

We see Smith's certitude and Haberler's spillover immediately in Figure 2. Smith's certitude shows up as Partner gains (the areas $a+b$) that result from the higher exports and higher border price. Since the FTA is reciprocal and nations are symmetric, Home gains the same from a higher border price and greater exports to Partner in good 2. Haberler's spillover shows up as the RoW lose (the area 'e') from the drop in the border price it faces (from $P-T$ to $P'-T$) and the reduction of its exports to Home.

The preference rent. A critical observation, as far as the regionalism-multilateralism debate is concerned, touches on a decomposition of Smith's certitude, namely how FTA exporters gain from two distinct features of their improved market access. First, the removal of the intra-FTA tariff boosts their market access directly. Second, FTA-based exporters benefit from the reduction in RoW exports induced by the tariff discrimination. The second part of the gain – area 'a' in Figure 2 – could be called the 'preference rent' since if the tariff cutting were multilateral instead of preferential, FTA partners would gain only 'b', not $a+b$. This preference rent 'a' is vulnerable to so-called preference erosion and as such, it plays a leading role in the stumbling bloc logic.

On the import side (Figure 3), Home gains a trade-volume effect (equal to area A) from expanding its imports, i.e. replacing high cost domestic production with lower cost imports. Home also gains from a border-price effect, i.e. the terms-of-trade improvement against RoW (area B) while losing from the terms-of-trade loss against Partner (area C_1+C_2). Home's terms-of-trade gain on the export side partly offset the terms-of-trade loss on the import side ($D_1=C_1$), so Home's net welfare change is $A+B+D_2-C_2$.¹³ As drawn it looks like Home and Partner gain, but this depends upon elasticities and the initial MFN tariff; in general Viner's ambiguity holds in this framework.¹⁴

The net welfare impact on RoW is unambiguously negative (Haberler's spillover). RoW experiences no change on the import side, but twice loses area 'e' (rightmost panel in Figure 2) – once on its exports of

¹³ Area C_2 might be called the 'trade diversion' effect, while D_2 and A might be call the 'trade creation' effect but as usual the trade creation/diversion dichotomy is incomplete; here it leaves out the third nation terms of trade gain, B.

¹⁴ As we shall see in the mathematical version of these diagrams, the FTA lowers welfare when the MFN tariff is sufficiently high. See Box 4.

good 1 to Home and once on its exports of good 2 to Partner. The ‘Haberler spillover’ is an externality as far as the global trade system is concerned and as such it plays a central role in the big-think regionalism literature.

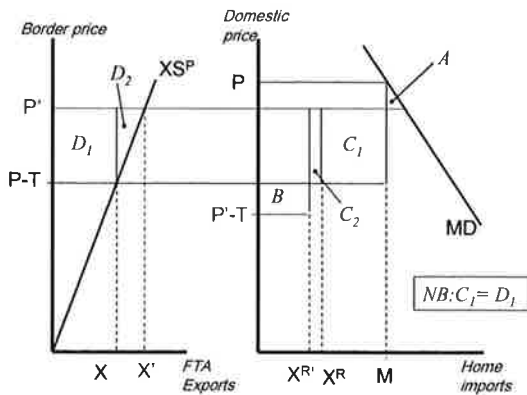


Figure 3: Ambiguous net welfare effects

2.4. Influential diagrams that ignore some of the 3 elemental effect

Up to 1990s, the literature’s main points were presented using diagrammatic analysis – two diagrams were particularly pivotal. The fact that these diagrams ignored some of the three basic effects distorted the direction of the literature and with it academic trade economists’ perceptions of RTAs. Since these older, incomplete diagrams occasionally enter today’s regionalism debate, it is worth presenting them briefly and highlighting their shortcomings. The first is the Johnson diagram that is still used in most undergraduate textbooks.

2.4.1. The Johnson (1960) diagram and the JCM proposition

Meade’s analysis was not integrated into mainstream trade theory in part because it was marginal and trade economists were interested in studying the discrete liberalisation implied by RTAs. Viner provided no diagrams so “Customs union theory”, as it was known at the time, was a distinctly wordy subject until Johnson (1960) introduced his famous diagram that illustrated Viner’s ambiguity with in a manner that was immediately transparent to all economists.

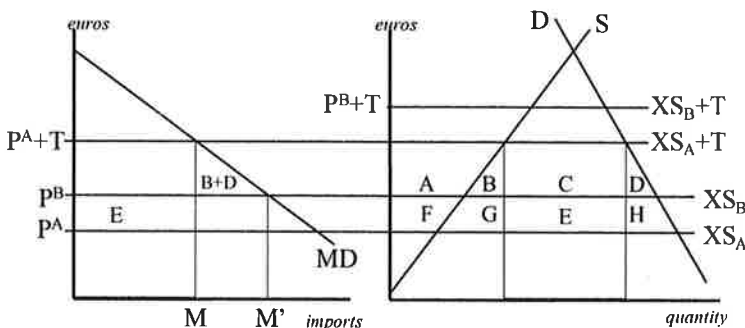


Figure 4: The Johnson diagram (small Home and Partner nations)

Source: Johnson (1960a) for right panel; left panel is a trivial transcription in import-price space.

Note: The right panel is the standard open-economy supply and demand diagram in price-quantity space for an infinitely small nation (Home). The left panel transcribes the analysis into a more compact diagram into price-import space.

In the diagram, Home imports can come from partner nation A or B. Home's demand is an infinitely small share of world demand, so it faces perfectly elastic export supply curves from both sources (labelled XS_A and XS_B). We start with the Home imposing an MFN specific tariff of T , so all imports come from the low-cost supplier, nation A. The domestic price is P^A+T while the border price is P^A .

Home can form a customs union with nation A or B so we consider both. The customs union with B would remove the tariff only on imports from B (the high cost supplier) and this produces supply-switching. Home switches from importing everything from A to importing everything from B. Home's domestic price falls from P^A+T to P^B . Assuming a utilitarian metric, the net welfare effects are $(B+D)$ minus E which may be negative or positive depending upon elasticities and the height of the initial tariff; this is Viner's ambiguity.¹⁵

The customs union with nation B was called "purely trade diverting", yet if the initial tariff is high and the P^B-P^A border-price gap was small, it can be welfare improving for Home. This result – a welfare-improving but purely trade diverting customs union – seemed to contradict Viner's reasoning and it produced the first of what was to be a long series of ivory-tower debates over terminology; this one pitted Meade (1955b) against Johnson (1960a) and Corden (1965).¹⁶

If Home chooses to form a customs union with A, the ambiguity disappears. Such a customs union is unambiguously welfare improving since its positive effects are identical to MFN free trade (both before and after all imports would come from A). Home's domestic price falls from P^A+T to P^A and the net welfare gain is $B+G+D+H$.¹⁷

Omitted elemental effects. Readers will immediately note that Smith's certitude and Haberler's spillover are missing. Third nations are entirely unaffected by the trade policy of an infinitely small nation like Home. In Johnson's diagram, the partner nations care no more about Home's trade policy than a perfectly competitive firm does about gaining or losing one atomistic buyer. This was an attractive feature in the Small-Think regionalism literature where national welfare was the pivotal issue, but it renders the diagram useless for the Big-Think regionalism debate. Quite simply, the diagram assumes that Home's decision to form an FTA has no systemic effects at all. Also missing from the diagram is an analysis the preferential access that Home's exports win in its partner's market.

For two decades, the Johnson-diagram dominated economic analyses of RTAs to such an extent that Smith's certitude and Haberler's spillover came to be largely forgotten by academic trade economists. This went so far that many mainstream trade theorists came to view RTAs as economically irrational – a view encapsulated in the Johnson-Cooper-Massell (JCM) proposition which states that a small

¹⁵ The left panel translated the effects into Meade's two-part framework: $B+D$ is the trade volume effect (related to the change in the volume of imports) and E is the trade price effect (related to the change in the border price).

¹⁶ The 1950s, 1960s and 1970s saw a rather extended and fruitless discussion of what Viner really meant. It featured contributions from the greatest trade economists of the time including Meade (1955), Johnson (1960a), Corden (1965), Bhagwati (1971, 1973), Kirman (1973) and Johnson (1974). Viner himself participated in the exchange without fully clarifying matters (Viner 1965). See Kowalczyk (1992) and Pomfret (1986) for summaries of the 'what Viner really meant' literature.

The quandary was thickened by the fact that the deep economics of taxation in a Walrasian world really does only need two effects, as Meade (1955) demonstrated, but only one of the two deals with trade volumes. One reaction to this cognitive dissonance was to expand the terminology (e.g. external trade creation, gross trade creation, etc. as in Balassa 1967) another was to stretch the meaning of the terms to cover the trade price effects (which is possible since bilateral border prices and imports are related by the export supply curves).

¹⁷ This contrast is the source of the rule of thumb that an FTA with your main trading partners is more likely to be welfare improving since you are giving preferences to the partners that have demonstrated themselves to be the low cost supplier by winning the largest market share in an even competition with other suppliers.

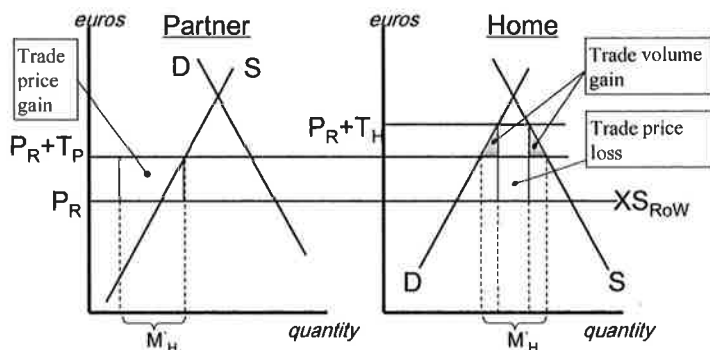
nation should always prefer unilateral MFN liberalisation to any RTA. The point is easily illustrated in Figure 4; cutting T to zero on imports from A and B would always yield net welfare gains that are at least as high as any customs union.

From the modern perspective, Johnson's analysis seems impossibly simplistic and the disconnect between academic and real-world thinking is truly astounding. For instance, when Britain put in its first application for EU membership in 1961, better market access for British exporters was the key concern, but academic economists were working with the Johnson diagram that assumed this away. Moreover, the main RTA in existence at the time – the EEC – counted for a substantial fraction of world imports and the key nations – Britain, France and Germany – were far from atomistic. As Pomfret (1997) points out, a number of frameworks were developed at the time that would have allowed the necessary extension, including Johnson (1957, 1958a), Humprey and Ferguson (1960), and Blackhurst (1972), but Johnson diagram's hold on the literature was so firm that the early efforts were obliged to stick with his small-country fiction.¹⁸

2.4.2. The "Small FTA" diagram

An important analytic extension of the Johnson diagram came with the 'small PTA' analysis (Shibata 1967). It allows for Smith's certitude even though it continues to assume away Haberler's spillover. The diagram continues to be used even today (e.g. Grossman and Helpman 1995), so it is worth presenting it briefly.

The small FTA diagram looks somewhat different under various assumptions on the pattern of comparative advantage and the smallness of the two partners. The various combinations of assumptions yield a range of results and that have been covered by three decades of literature (see Panagariya 1999 for a comprehensive survey of papers using the small PTA diagram in recent decades). Here we study a fairly standard case and illustrate the diagram's properties by demonstrating two classic results in the



regionalism literature.

Figure 5: The small PTA diagram, a simple case

Source: Author's modification of a diagram in Pomfret (1997).

The diagram presumes that the two FTA partners, Home and Partner, import the same good from the rest of the world (RoW). Home and Partner are "small" with respect to RoW and so face a perfectly elastic RoW export supply curve, $X S_{RoW}$. This sets the initial border price to P_R in both nations. Home

¹⁸ The early 1980s saw a number of widely read studies that sought to reverse the JCM proposition while staying in the small country framework. These efforts, e.g. Wonnacott and Wonnacott (1981) and Berglas (1983), strike the modern reader as awkward due to the small nation assumption and the intricate diagrammatic analysis.

has a higher MFN tariff than Partner to start with – T_H as opposed to T_P – so the pre-FTA price is higher in Home.

When Home and Partner form their FTA, Partner-based firms initially see a higher price in Home and so begin exporting to Home. In equilibrium, all post-FTA Home imports, M'_H , are supplied by Partner firms. Partner's internal price remains at P_R+T_P , so its consumption and production are unchanged, which means that the new exports to Home are replaced – one for one – by new Partner imports from RoW. In the case illustrated in the diagram, Partner is large enough relative to Home to ensure that Home's entire demand can be satisfied by Partner producers at P_R+T_P . In terms of welfare, the FTA results in a positive trade volume effect for Home but a negative border price effect (Home pays P_R+T_P for its imports instead of P_R). Partner expands its imports across the tariff wedge and this results in a positive trade volume effect equal to T_R times the expanded imports, i.e. M'_H .

Irrelevance of rules of origin. Although it seems an odd objective from today's perspective, Shabata's goal was to illustrate the irrelevance of rules of origin (ROOs). His point is that the ROOs only prevent blatant trade deflection. Because goods from Partner and RoW are fungible, the equilibrium is the same with and without rules of origin as long as Partner's supply is sufficient. If Partner's supply were not large enough to supply all of Home's imports at P_R+T_P , the FTA with ROOs would have somewhat higher prices than one without. The use of this exercise in the Big-Think regionalism literature comes in the form of 'imported liberalisation/protection' (Section 3.3.4), and Grossman and Helpman (1995).

Un-sustainability of the FTAs. Another application that found popularity in the academic literature but seems odd today is the proposition that FTAs will always breakdown. Using a diagram like Figure 5, Vousden (1990 p. 234) argues that Home would be tempted to lower its MFN tariff to just under that of Partner in order to recapture the tariff revenue and Partner would have an incentive to reply. The resulting race-to-the-bottom tariff cutting was viewed as making FTAs 'unsustainable.' Vousden (1990) did not attract much attention until Richardson (1995) extended and popularised it. These two results (irrelevance of rules of origin and unsustainability of FTAs) are classic examples of how academic thinking on regionalism has often followed literature-driven paths that have little relevance to real-world policy concerns.

3. STUMBLING AND BUILDING BLOC LOGIC

From 1960 to the late 1980s, regionalism was a simple matter. It consisted of: 1) the EEC which encompassed a third of world trade in a highly effective customs union, and 2) a slew of RTAs among developing nations that covered a trivial fraction of world trade and in any case never operated effectively. Regionalism's systemic implications were simply not an issue.

Regionalism got complicated in the late 1980s when Canada and Mexico changed their minds on regionalism (Krugman 1991b p.7).¹⁹ The US had long been interested in regional preferential trade, but

¹⁹ Bhagwati (1991 p.71) ascribes the shift to the US's conversion to regionalism, but this contradicts the judgments of trade policy scholars who were engaged in the details of policy at the time (Smith 1988 p.41; Wonnacott 1987 p.17; Schott 1988 p.29; Hufbauer, Schott and Clark 1994 p.100; Whalley 1993). It also contradicts the facts. The US's long-standing interest in regionalism is testified by a long string deals that were struck, or almost struck in 1854, 1874, and 1911. In March 1948, they concluded a secret draft protocol eliminating most tariffs and quotas bilaterally, but this was ultimately rejected by the Canadians. In 1958, US government procurement was preferentially liberalised in Canada's favour. The US-Canada Auto Pact came into force in 1965. The 1974 Trade Act authorised the US President to negotiate an FTA with Canada, and the 1979 Trade Agreements Act required the President to study an FTA in North America.

Mexico and Canada, by contrast, had always resisted North American regionalism, fearing domination by their giant neighbour. Canada overcame its traditional resistance to propose an FTA in 1985 (it entered into force in 1989). Mexico

Mexico and Canada resisted, fearing domination by their giant neighbour. Canada propose an FTA with the US in 1985 that entered into force in 1989. Mexico proposed an FTA with the US in 1990 and this evolved into NAFTA at Canada's insistence (to safeguard its Auto Pact preferences). The US-Mexico initiative triggered a wave of Latin American requests for bilateral FTAs with the US and gave greater urgency to arrangements among Latin Americans most notably Mercosur.²⁰

The rise of North American regionalism coincided with two other major development in the world trade system. First, GATT negotiations were lurching from crisis to crisis in the late 1980s and then, as mentioned in the introduction, seemed to die with the acrimonious collapse of the Uruguay Round's 'final' summit in December 1990. Second, European regionalism was reignited by the Single European Act and the collapse of the Soviet Union.

Many respected thinkers looked at this temporal correlation and saw causality. They feared that regionalism's spread might kill the proverbial gold-laying goose – the GATT-centred world trade system. The fears are easy to understand. Two-thirds of world imports went to North America and Europe; 40% of this was intra-bloc trade and soon to be covered by discriminatory liberalisation schemes. Still more worrisome, North American and Europeans were the stalwarts of the GATT system. If regionalism weakened their support of multilateralism, the goose was indeed in deep trouble. Spreading regionalism had become much more than a small-think "should I join?" question.

These fears promoted regionalism to a status on the world's policy agenda that it had not enjoyed since the 1950s. This naturally attracted paradigm-setting efforts from the profession's leading international economists.

3.1. Framing the new regionalism debate

Krugman (1991b) is clearest in rejecting the relevance of the 1950s small-think approach and delineating the outlines of a new line of inquiry – what I call Big-Think Regionalism:

“In a fundamental sense, the issue of the desirability of free trade areas is a question of political economy rather than of economics proper. While one could argue against the formation of free trade areas purely on the grounds that they might produce trade diversion ... [t]he real objection is a political judgment: fear that regional deals will undermine the delicate balance of interests that supports the GATT.” (Krugman 1991a, p.23)

The Krugman (1991b) framing of what he identified as the key issue – the impact of regionalism on support for the GATT system – did not catch on.²¹ His 1991 papers, however, did re-frame the 1950s

overcame its traditional fears and proposed a bilateral FTA with the US in 1990. The US accepted immediately and the bilateral evolved into NAFTA at the insistence of Canada (whose main interest was safeguarding its Auto Pact preferences). The US-Mexico initiative triggered a wave of Latin American requests for bilateral FTAs with the US and gave greater urgency to arrangements among Latin Americans most notably Mercosur. See Baldwin (1997) or Serra et al. (1997) for an account of this domino effect.

²⁰ See Baldwin (1997) or Serra et al. (1997) for an account of this domino effect.

²¹ He argued that the multilateral process had run aground with the December 1990 failure and was unlikely to get afloat anytime soon as the system was plagued by deep-seated problems. “[W]hile some kind of face-saving document will probably be produced, in reality the Uruguay Round has clearly failed either to significantly liberalize trade or to generate good will that would help sustain further rounds of negotiation.” Regionalism, however, was not one of those deep-seated problems. “But while the move to free trade areas has surely done the multilateral process some harm, it is almost surely more a symptom than a cause of the decline of the GATT. ... [T]he problems of the GATT are so deep-seated that it is unlikely that a world without regional free trade agreements would do much better.” He closes his essay with a prediction that history falsified and *faute-de-mieux* view on regionalism. “The world may well be breaking up into three trading blocs; trade within those blocs will be quite free, while trade between the blocs will at best be no freer than it is now and may well

national welfare question into a global-level question. Krugman (1991a) introduced a new approach by asking whether spreading regionalism raises or lowers world welfare. This spawned a decade-long literature and continues to influence research even today.

This “Is bilateralism bad?” literature – also known as the multilateralism versus regionalism literature – looks distinctly odd from today’s perspective. It tries to use simple theory to answer what is intrinsically a complicated empirical question. At the time, however, it was the best they could do. Economists had limited access to the necessary data and lacked the panel econometric techniques to exploit them. Moreover, spreading regionalism was at the time more of a threat than a reality, so there was little to test empirically. In my mind, this literature is now mainly of interest to historians of thought, but many participants in today’s regionalism debate cite it, so I review it in Section 3.5.

The focus of this section is on what I consider to be the central theoretical question: Does regionalism help or hinder multilateralism? Ultimately this also is an empirical question, but given the relative little experience the world has had the regionalism-multilateralism interface (only one MTN has been completed since 1991), convincing empirics is not yet feasible – although some tantalising results are beginning to emerge. Moreover, given the complexity of the inter-linkages, a clear theoretical understanding is necessary condition for well-structured empirical work.

3.1.1. Are regionalism and multilateralism friends or foes?

Bhagwati’s book, *The World Trading System at Risk*, does not focus on regionalism. Indeed, his Part One, “The GATT Architecture under Threat” listed four main threats; regionalism was number four. Nevertheless, his writing helped establish Big-Think Regionalism as the new paradigm. In the lead paragraph to his chapter on regionalism he writes: “These regional alignments have led to fears of fragmentation of the world economic into trading blocs in antithesis to GATT-wide multilateral free trade. Does such regionalism truly constitute a threat to multilateralism?” Although he does not set out an analytic framework for answering the question, his writing influenced the intellectual paradigm for more than a decade.

Framing the issue three ways

Theory requires explicit questions. Asking whether regionalism and multilateralism are friends or foes is not sufficient. Pure logic identifies three mutually compatible ways that regionalism and multilateralism could interact:

- Regionalism could affect multilateralism,
- multilateralism could affect regionalism, and
- both multilateralism and regionalism could be driven by third factors.

The literature has looked at all of these, but the first has dominated since Krugman (1991b, 1993) presented a simple analytic framework for posing the question. His explicit question was: how does an exogenous variation in regionalism (specifically, the formation of a new RTA) affects nations’ incentives to cut tariffs multilaterally?

Modelling choices and branches of the literature. Answering this question requires an economic model that links tariff choices to equilibrium outcomes and a political economy model that endogenises the MFN tariff choice (the RTA tariff levels are taken as exogenous in this literature). The choice of economic models is quite open, but the literature naturally gravitated to the simplest possible models

be considerably less free. This is not what we might have hoped for. But the situation would not be better, and could easily have been worse, had the great free trade agreements of recent years never happened.”

that yielded the elemental economics effects – Smith’s certitude and Haberler’s spillover – that are pivotal to regionalism’s systemic implications (Viner’s ambiguity is more a Small-Think regionalism issue).

The choice of political economy modelling is facilitated by two key institutional features of the global trade system: the nature of RTA tariff cutting and MFN tariff cutting. Specifically:

- Real-world RTAs involve bargaining on very few tariff lines; they cut tariffs to zero on most goods.²²

This institutional fact made modelling of the preferential liberalisation easy – most authors assumed that forming an RTA meant zero tariffs on all good traded among members.

- Real-world multilateral tariff-cutting talks also involve bargaining on very few tariff lines.

Since the 1963 Kennedy Round began, multilateral tariff ‘bargaining’ was simple: GATT members pre-agreed tariff-cutting rules with exceptions for specific products (the usual ‘sensitive’ products such as clothing, footwear, etc.). This real-world feature is easily understood. Nations’ tariff schedules typically list about ten thousand individual lines and at least a dozen nations participate actively in MTN tariff-cutting negotiations. (The number has increased greatly in the on-going Doha Round, but only the richest OECD nations participated substantially in the Kennedy, Tokyo and Uruguay Round tariff cuts.) If each of the dozen nations bargained for one minute over each of its tariff lines with each of the other nations, the talks would take 25 years of 24/7 discussion. Trade diplomats avoid this by agreeing the basic tariff cutting before the talks start – i.e. it is set in the agenda. Specifically, the Kennedy and Tokyo Rounds used formulas that cut tariffs by about one-third with some sensitive industrial goods excluded (Winters 1991 p. 171). The Uruguay Round agreed, at its 1988 Ministerial Midterm Review in Montreal, to cut tariffs by at least as much as in the Tokyo Round, i.e. by about one-third and again sensitive products were predictably excluded (Croome 1995, p.183).

This institutional fact made modelling of the multilateral liberalisation easy (or at least should have). Many authors modelled multilateralism as saying yes-or-no to an exogenously defined MFN tariff cut. One branch of the literature however adopted the alternative tack of twisting reality to the theory. These authors ignored the institutional features of MTNs in order to use game theory’s simple bargaining models – models that had been developed to deal with much simpler bargaining situations. While I do not believe that such models provide useful insights to the key Big-Think regionalism questions, they are often referred to in the academic literature, so I review them in Section 3.4.

While most of the literature has followed Krugman’s lead in asking how exogenous variations in regionalism affect multilateralism, two other branches consider how 1) RTAs can affect MTNs (competitive liberalisation), and 2) How RTAs and MTNs are driven by deeper forces (Summers’ notion that all the ‘isms are good). I review thinking on these in Section 3.5.

With this structuring of the literature out of the way, I turn to review what I consider the most relevant theory on the Big-Think Regionalism literature.

3.2. Stumbling bloc logic

In its cleanest form, the stumbling bloc logic asserts that if the stumbling bloc RTAs were forbidden, global free trade would be obtained, but since they are allowed global free trade becomes impossible.

²² RTAs involving a developed WTO member must respect this; those involving only developing nations do not, but de facto they have as the chapter in this volume “Market access provisions in regional trade agreements” shows.

Weaker forms are put forth in the help-or-hinder literature – for example RTAs may slow the achievement of global free trade – but clarity has led the profession to focus on the strong form.

In my opinion, only three distinct stumbling-blocs logic are relevant to real-world policy analysis: 1) the preference-erosion/exploitation stumbling bloc, 2) the goodies-bag stumbling bloc, and 3) the cherry-picking stumbling bloc, although I am sure many more shall be illuminated in coming years. For simplicity’s sake, these stumbling-blocs logics are demonstrated under the naïve but transparent assumption that national governments choose tariffs to maximise national welfare.

3.2.1. Preference-erosion/exploitation stumbling bloc

The logic of the preference-erosion/exploitation stumbling bloc is dead easy and quite general. Starting from a world where all nations have MFN tariffs, the question is: Can some group of nations raise their collective welfare above the free trade level by forming a trade bloc and thus exploiting other nations? If the answer is “yes,” then that bloc is a stumbling bloc on the road to multilateral free trade because the bloc members would veto global free trade as undermining their exploitation of third nations.

In the Walrasian setting, and in many other trade models, the answer is almost always “yes”, but the answer may depend upon the level of MFN tariffs when the bloc is formed. Given Smith’s certitude and Haberler’s spillover, some combination of nations is bound to be able to better exploit third nations by acting as a bloc. This is almost trivially true if the bloc can violate its WTO tariff-bindings by raising external tariffs. After all, the block as a whole has more buying power than its constituents do individually, so the bloc can better exploit foreigners. Less obviously, but equally true, stumbling blocs can be found even when external tariffs are maintained (as has been the case for all of the major post-war RTAs).

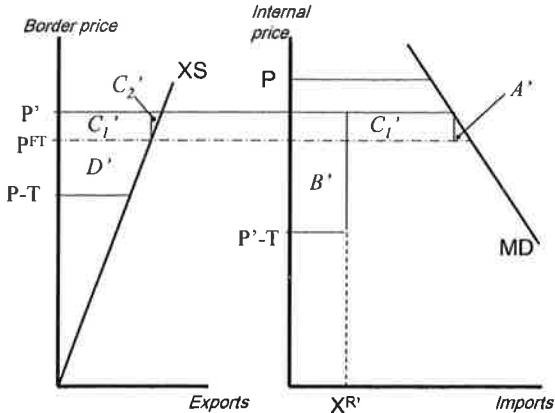


Figure 6: Net welfare effects, FTA to global free trade

The preference-erosion/exploitation stumbling bloc logic can be illustrated more concretely with the RTA diagram.

An Walrasian illustration

Start by noting that MFN free trade would be approved by all governments in the simple model laid out above (Figure 2 and Box 4). The first welfare theorem tells us that global free trade is efficiency enhancing (a move to the first best) and symmetry ensures that each nation would get an equal slice of the gains. This conclusion, however, can be reversed when we start from the situation where Home-Partner have formed an FTA. Taking the FTA as the base case, a move to global free trade eliminates the preference margin Home exporters enjoyed in Partner and this would lead to a terms-of-trade loss of area C_1' and a trade-volume loss of C_2' (see Figure 6). On the import side, global free trade would

win Home an additional trade volume gain of area A' , a terms-of-trade improvement with respect to Partner exporters of area C_1' , and a terms-of-trade loss on imports from RoW, area B' . Global free trade would also improve Home exporters' market access to RoW and this would boost Home welfare by area D' . Overall, the net welfare change of moving from the FTA to global free trade is $-C_2' - B' + D' + A'$; the sign of this is ambiguous. As Box 4 shows, however, it is always negative when the initial MFN tariff is low enough.

Working out magnitudes in the diagrams for the whole range of possible T 's is tedious, so we lighten the analysis by introducing the equations behind the diagrams (see Box 4). Solving for equilibrium prices under the FTA and global free trade situations and plugging these into the welfare functions, we can plot the welfare of a typical FTA nation (Home) against the MFN tariff. The results are shown in Figure 7.

For initial MFN tariff levels that are sufficiently low, we see that Home's welfare is higher with the FTA than it is with global free – and this despite the fact that Home would have agreed to global free trade starting from the initial situation without the FTA. (The line marked MFN is everywhere below the line marked global free trade, but the dashed FTA line is above the global free trade line for sufficiently low tariffs.)

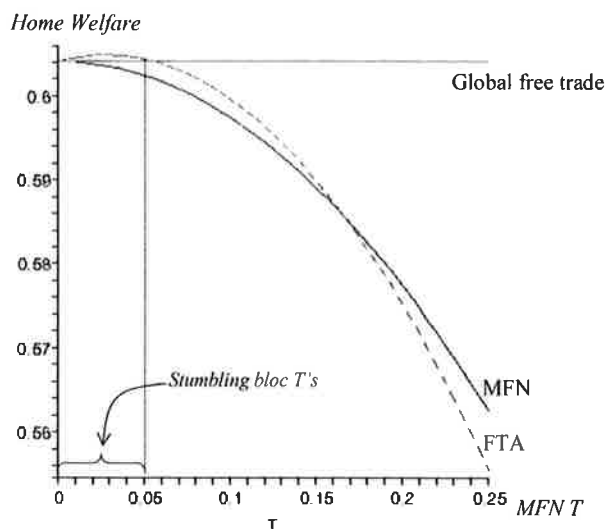


Figure 7: Stumbling bloc FTAs

Notes: Home welfare is plotted against the MFN tariff for 3 trade regimes; MFN = all nations impose same MFN tariff T on all goods, FTA = Home and Partner form an FTA, and global free trade which is the MFN regime with $T=0$. The parameters chosen are $a=1$, $b=1/2$; the qualitative results are unaltered for other choices of parameters. Note that $T=1/4$ is the prohibitive tariff with $a=1$, $b=1/2$. In ad valorem terms, the Viner crossing occurs at a tariff of about 27% of the free trade price, while the ad-valorem tariff threshold for a stumbling bloc occurs at about 9%; the probative tariff is about 42% for these parameter values.

Intuitively, the FTA allows Home to exploit RoW both on the import side (by pushing down the price it pays RoW good-1 exporters) and on the export side (by raising the price in Partner at the expense of RoW good-2 exports). The move to global free trade undoes these two forms of exploitation, but in exchange it provides better access to the RoW market and more liberalisation in Home's import market. When the initial T is low, the market-access and home-liberalisation gains are modest so the net is negative. In other words, the basic logic of the stumbling bloc result turns on the way that an FTA allows the FTA partners to exploit excluded nations.

The Walrasian model presented here is very special, but the heart-and-soul of this stumbling-bloc effect – the exploitation of excluded nations – is quite a general result; one that is surely an important consideration in the real world.

The opposition of small developing nations to agricultural liberalisation in the DDA – especially those who benefit from the EU’s unilateral preferences – is a classic example of the preference-erosion stumbling bloc. Had the EU not unilaterally granted these nations preferences, these nations would probably have been pushing for EU market opening in sugar and other goods.

Who did what when? The logic of the preference-erosion/exploitation stumbling bloc was demonstrated in a Walrasian setting by Reizman (1985), Kennan and Reizman (1990), and in a Brander-Krugman setting by Krishna (1998) and Freund (2000a, b).²³ The core algebra for the preference-erosion/exploitation stumbling bloc in the Brander-Krugman setting is laid out in Box 6 where the main Freund and Krishna results are also demonstrated.

3.2.2. Goodies-bag stumbling bloc

Another stumbling-bloc logic – one which follows closely the fundamental economic logic of the preference-erosion stumbling bloc – might be called the ‘goodies bag’ stumbling bloc.²⁴ In a nutshell, the rents corresponding to ‘Smith’s Certitude’ can be thought of as a ‘bag of goodies.’ These goodies can be used by one or both RTA parties to buy non-economic benefits from its partners. Since the size of non-economic benefits that can be ‘purchased’ is linked to the richness of the ‘goodies bag’, i.e. the margin of preferences, RTA members have an extra incentive to maintain high margins of preference by avoiding multilateral liberalisation. The goodies-bag logic, however, extends to a far greater range of issues than the tariffs that are the focus of the preference-erosion stumbling bloc. In the case of RTA between very large and very small nations – a case that is extremely common in the new century (e.g. US and Costa Rica or Japan and Singapore) – the large country’s interest in the RTA can hardly be the preferential market access.

The EU, for example, grants extensive preferences to its members’ former colonies, justifying these on the basis of international solidarity. In other words, the economic gains to the EU’s partners count as a plus inside the EU since they advance one of the EU’s non-economic objectives – fostering development. Similarly, but more explicitly, the US justifies many of its FTAs with small, poor nations on the basis of non-economic objectives (typically anti-drug and/or anti-terror policies).

The previous section illustrated how the desire to guard rents created by an RTA could make a nation reject global free trade when MFN free trade would have been embraced without the RTA. The goodies-bag stumbling bloc logic amplifies this mechanism by making both nations interested in each other’s export rents – the area corresponding to C_1 ’ in Figure 6, with the link coming through the pursuit of non-economic objective (non-economic in the narrow sense).

Who did what when? The theoretical notion was formalised by Limão (2006) and tested in Karacaovali and Limão (2005) for the EU, and Limão (2006) for the US. Both papers suggest that these trade powerhouses cut their MFN tariffs less for items where there granted important preference

²³ Also see Goto and Hamada (1995a, b), Nordstrom (1995) and section 3 of Bond and Syropoulos (1996a) for example of preference erosion/exploitation stumbling blocs.

²⁴ Here is an example of the terminology in the context of US unilateral concessions from The Hindu newspaper, 24 September 2001. Referring to the mood in Washington, the correspondent note: “... There is tremendous support and sympathy for Pakistan in the light of the decisions taken by its President, Gen. Pervez Musharraf, to support the U.S. wholeheartedly in its war against terrorism. As for Pakistan, the question is not merely the lifting of U.S. sanctions. Rather, it is in the larger “goodies bag” that most certainly is coming up. As it is, there is disappointment in the official Pakistan establishment here that Mr. Bush has not gone the whole hog.” <http://www.hinduonnet.com/2001/09/25/stories/03250007.htm>.

margins. In short, the goodies-bag did seem to act as a ‘slowing bloc’ even if it did not act as a stumbling bloc.

The evidence in Limão (2006) is widely misrepresented as showing that the US raised tariffs in the Uruguay Round for items on which it granted FTA preferences. Of course, this cannot be right since MTN market access talks only involve tariff bindings and the US did not violate any of its bindings in the Uruguay Round. Indeed, the data shows US tariffs decreasing for all but 12 of the thousands of tariff lines (defined at the HS-8 products level in the WTO’s database).²⁵ Formally, Limão estimates an econometric model of US tariff cuts in during the Uruguay Round. His famous stumbling bloc results is that he finds is that the US cut tariffs by less than his econometric model predicted they should have on items for which the US had granted FTA preferences before the Uruguay Round. In short, he shows that the US preferences acted as a ‘slowing block’ not a stumbling bloc.

3.2.3. Cherry-picking stumbling bloc

An entirely distinct mechanism is at work in the third type of stumbling bloc – the cherry-picking stumbling bloc. Moving to global free trade will typically involve some pluses and some minus from the national perspective. The question is: Can one find a group of nations whose integration would provide many of the pluses with few of the minuses? If the answer is “yes,” the trade bloc is likely to be a stumbling bloc. Starting from the bloc situation, a move to global free trade involves many minuses and fewer pluses for bloc members, so bloc members may veto global free trade even when they would have embraced it without the trade bloc.

Here is an example of a setting that could yield a cherry-picking stumbling bloc. Assume the trading environment is marked by Helpman-Krugman trade; there is both intra-industry trade in differentiated products and inter-industry trade in Walrasian sectors. In this world, trade liberalisation will produce gains from trade due to the Krugman variety effect. It will also produce gains from trade from comparative advantage effects, but the comparative advantage gains come bundled with politically difficult Stolper-Samuelson effects on domestic factor prices. Now suppose two large nations have quite similar factor endowments. If they form a trade bloc, they will win a large share of the variety gains that would come with global free trade, but will experience little of the Stolper-Samuelson pain. Taking the trade bloc as the base case, the bloc members may find a move to global free trade unattractive. It would entail a good deal of Stolper-Samuelson pain and only a modest amount of additional variety or comparative advantage gains. Depending upon parameters, especially the political power of Stolper-Samuelson sufferers, the gains may not be sufficient to make global free trade attractive to the bloc members.

Who did what when? Levy (1997) illustrates a cherry-picking stumbling bloc in a highly stylised setup, but his main result is surely more general than his model.

3.3. Building bloc logic

While many trade policy scholars – such as Paul Krugman (1991b) and Jagdish Bhagwati (1991) – worried that regionalism was a stumbling block to global free trade, others – such as Larry Summers (1991) and Fred Bergsten (1991) – viewed regionalism as a largely benign or even as a constructive force in the world trade system.

²⁵ As per a bilateral communication with the author.

This sub-section consider the economic logic of the assertion that RTAs could foster multilateral liberalisation, i.e. how RTA may be building blocs on the road to global duty-free trade. There are four main logics in the literature. We start with the one that permeates the justifications used by nations which simultaneously pursue regional and multilateral liberalisation – the notion that preferential liberalisation creates a political-economy momentum that makes multilateralisation liberalisation easier (and vice versa).

3.3.1. Juggernaut building bloc logic

Liberalisation begets liberalisation according to the juggernaut building-bloc logic. The logic comes in two parts that are most easily explained in the context of multilateral liberalisation. When the GATT started in 1947, import duties were high worldwide since they had been set without international coordination during the Smoot-Hawley tariff wars. The tariffs balanced the supply and demand for protection in the ‘political market’ of each nation separately. The main demanders of import protection were import-competing firms and workers they employed. The governments were the supplier of protection but concern for the general economic wellbeing of the nation meant that the supply of protection was not perfectly elastic.

Starting from this situation of uncoordinated tariff-setting, announcement of an MTN based on the principle of reciprocity alters the array of political forces inside every participating nation. The central point is reciprocity; it converts each nation’s exporters from bystanders in the tariff debate to anti-protectionists. For exporters, lobbying against domestic tariffs becomes a means of lowering foreign tariffs. Because the MTN rearranges the political economy forces inside every nation, all governments find it politically optimal to choose tariff levels that are lower than the unilaterally optimal tariffs.²⁶ This is the first part of the juggernaut theory.²⁷ The logic is not new.

Informed observers have long known that the GATT’s reciprocal MTNs was mostly about helping nations internalise a political economy externality inside their own polities, i.e. about making it easier for national politicians to put together a national coalition in support of freer trade. Writers such as Robert Baldwin (1970, 1985) and Mac Destler (1986) are explicit on this point, but historical accounts of the Cobden-Chevalier Treaties show that using external trade deal to re-align domestic political forces was very much in the minds of 19th Century thinkers (Irwin 1993 p.96). Even Krugman (1991b) writes: “[T]he process of multilateral negotiation ... sets each country's exporting interests as a counterweight to import-competing interests; as trade negotiators bargain for access to each others' markets, they move toward free trade despite their disregard for the gains from trade as economists understand them.”

The second part of the juggernaut logic concerns the impact of the tariff cuts on openness. The tariff cuts make all nations’ more open – export sectors expand with the foreign tariffs cuts, and import-competing sectors contract with domestic tariff cuts. Assuming political influence is linked to industry size, this economic re-landscaping strengthens pro-liberalisation forces and weakens anti-liberalisation

²⁶ More formally, without MTNs governments maximize a politically weighted objective function that includes things affected by the nation’s own tariffs – profits of import-competing sectors, consumer surplus and tariff revenue. During the MTN, a nation’s tariff affects all these things, but foreign tariff levels are linked to domestic tariffs via reciprocity, the objective function now includes the impact of foreign tariffs on exporter’s profits. Since this new impact is negative (higher domestic tariffs lower exporter’s profits via reciprocity), announcement of the MTN leads the government to find it politically optimal to choose a tariff that is lower than the political optimal tariff before the MTN.

²⁷ The word ‘juggernaut’ – defined as “any massive inexorable force that advances crushing whatever is in the path” – stems from a British mispronunciation of the Hindu diety of the Puri shrine, Jagannath. A festival is held in Puri involving the ‘chariot of Jagannath’, an enormous and unwieldy construction that requires thousands of people to get it rolling. Once started, however, it rolls over anything in its path.

forces in all nation – although of course such industrial restructuring takes years. In other words, the initial reciprocal tariff cuts start a liberalisation juggernaut rolling. Due to the economic re-landscaping that occurs during the phase-in of the initial tariff cuts, all governments find it that their politically optimal tariff in the next MTN is below the levels that they found politically optimal during the previous MTN. These fresh tariff cuts continue the re-landscaping and the juggernaut continues to roll forward. Thus once the liberalisation juggernaut starts rolling, it crushes all tariffs in its path.²⁸

To the extent that regionalism can start the juggernaut rolling, RTAs can be building blocs – hence the name juggernaut-building-bloc logic.²⁹ The precise mechanism is a simple extension of the juggernaut logic. RTAs re-landscape members’ economies – making export sectors larger and import-competing sectors smaller. Thus the RTA can alter the member governments’ stance in MFNs, making it politically optimal to cut MFN tariffs to levels that would not have been politically optimal without the RTA. Of course, if a RTA results in higher external tariffs (as in the case of the EU’s agriculture tariffs), then a RTA can start the juggernaut rolling backward.

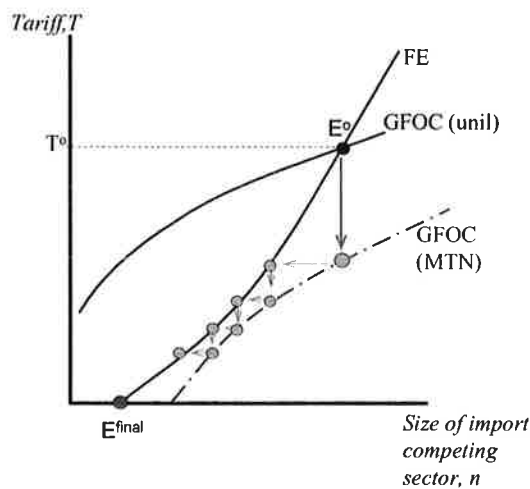


Figure 8: Juggernaut framework

Source: Baldwin (2006).

The basic idea can be presented in a diagram, Figure 8. The two curves, FE and GFOC, show how the size of the import competing sector depends upon the tariff (free entry) and how that tariff depended

²⁸ The juggernaut logic is from Baldwin (1994 chapter 2.5) and is elaborated in Baldwin (2005, 2006) and formalised in part by Freund (2000b). The first part of the juggernaut mechanism – reciprocity’s realignment of domestic special interests – has long been recognised in histories of trade liberalisation, e.g. Destler (1992 p.17) and Bergsten (1996) under the name “bicycle theory and export politics”; the point was also made by many others including Robert Baldwin (1984), and Hoekman, Bernard and Kostecki (2001). The basic idea dates much further back as Irwin (1996) points out. Even more recently, the first-half of the juggernaut logic has been studied formally by Grossman and Helpman (2001), and Bagwell and Staiger (2002). Juggernaut-like mechanisms were discussed independently by Hufbauer, Erb and Starr (1980), Hufbauer, Schott and Clark (1994, p.164), and Richardson (1993). Bergsten (1998) mentions an alternative source of political-economy momentum (“modest liberalization begets broader liberalization by demonstrating its payoff and familiarizing domestic politics with the issue”). Staiger (1996, section 5.4) uses a repeated game setting with workers moving slowing out of the import-competing sector to generate gradualism, but MTNs and GATT reciprocity play no role. Milner (1997) and Oye (1993), working independently in the International Political Economy context, discuss mechanism by which FTAs can create a pro-liberalisation political economy momentum. More recently, Hathaway, (1998) presents a similar logic in her positive-feedback model.

²⁹ But likewise, multilateral tariff cutting may lower tariffs to the level where PTAs become feasible, when previously they were not (this may have been the case with the Canada-US FTA).

upon the size of the import competing sector via politics. Here 'n' (the number of firms) measures the size of the import competing sector. The politically optimal tariff choice – taking as given the size of the import competing sector – is plotted as $GFOC_{unil}$ (solution to the government's first order condition without MTN). The politically optimal tariff rises with n since the larger is the import competing sector, the higher is the political benefit from a marginal increase in the tariff. The free entry curve – FE – relates the equilibrium number of firms to the tariff. As the tariff rises, more firms find it optimal to enter the market. These two relationships assume that the government and firms are short-sighted and the government chooses T taking n as given while firms choose n taking T as given. Note that this only captures the size of anti-trade forces; the size of the pro-trade export sector is suppressed to avoid a three dimension diagram.

To see the two steps of the juggernaut effect, note that announcement of the MTN shifts of the GFOC curve to $GFOC_{MTN}$. This is lower since the government finds it politically optimal to set a lower tariff for any given level of n when domestic protection is linked to foreign protection via reciprocity. As drawn, the new long run equilibrium, E^{final} , entails free trade, but since entry and exit occur slowly, the tariff and state of the import-competing industry do not jump to E^{final} . The figure illustrates one possible adjustment path. Each MTN results in an instantaneous drop in the tariff, but slow entry/exit means the horizontal arrows reflect slow movement of the state variable.

The juggernaut effect acts as a building bloc if the FTA reduces the importance of the import-competing industry in governments' objective functions. As we shall in the previous section, an FTA between Home and Partner does reduce the size of the Home import-competing sector, so in that model, FTAs are building blocs on the road to global free trade.

Figure 9 illustrates the juggernaut building bloc logic diagrammatically (the maths of the juggernaut stumbling bloc are illustrated in Box 7 for the Walrasian example). The FE curve in Figure 8 was drawn for symmetric MFN tariffs. When the FTA is signed, the FE curve rotates inward as shown since the additional competition from Partner producers lowers that Home price facing import-competing firms so some of them exit. Consequently, the GFOC under reciprocal trade trades will yield a lower MFN tariff after than FTA (point E^2) than before (point E^1). Of course, if the FTA somehow raised protection of the Home import-competing sector, the effect is reversed and the FTA acts as a stumbling bloc.³⁰

The Wei and Frankel momentum. Wei and Frankel (1995) illustrate another juggernaut-like mechanism. In their model, imperfect information makes workers uncertain as to whether they will win or lose from global free trade. Since an FTA is an intermediate form of liberalisation, they show that an FTA could be politically feasible even when global free trade would not. After the FTA is signed, the nation's true comparative advantage is revealed and workers now know whether they will win or lose from free trade. If the parameters are chosen carefully, the certainty resolution may mean that global free trade is politically feasible only after the FTA. Thus the FTA is a building bloc and since it operates by altering the political economy landscape it can be thought of as a momentum-generating mechanism.

³⁰ Agriculture in Europe is a good example. The customs union formation realigned special interests in the Community in a way that fostered higher agricultural tariffs. Since EEC tariffs on agricultural goods were not bound until the 1990s, the EEC was free to raise its agriculture tariffs without appealing to the Article 24 exception. Since EEC tariffs were bound for manufacturing goods, EEC nations need the Article 24 exception to establish the common external tariff (CET) and this led them to respect the Article's requirement that the CET not be higher on average; roughly speaking France lowered its tariffs, Germany raised its and the Benelux nations did little since their tariffs were initially between the French and German levels. In this way, the EEC customs union formation probable reduced the overall size of the EEC import-competing sector in manufacturing but raised it in agriculture.

The juggernaut logic exploits the fact that nations do have preferences over trade arrangements in the way that individual consumers have preferences over consumption bundles. The nature of the trade deal proposed can affect a nation's ranking over choices (unilateral versus reciprocal, for example). Moreover, nations' rankings over choices – unlike the rankings of a standard Walrasian consumer – are path-dependent since historical liberalisation can affect the current political strength of various pro- and anti-trade special interest groups. There are, however, a number of building-bloc logics that assume nations – like Walrasian consumers – have exogenous preferences over outcomes. The next section considers the easiest of these, assuming a representative consumer and assumes that the government acts to maximise this individual's wellbeing.

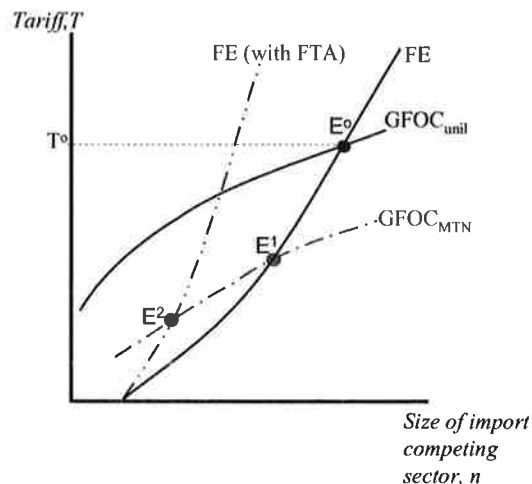


Figure 9: Juggernaut building bloc logic

3.3.2. Kemp-Wan building bloc logic

The assertion that trade blocs may be building blocs in a static Walrasian world is as easy and as general as the assertion that they may be stumbling bloc. Starting from a world where all nations have MFN tariffs, the question is: Can some group of nations always raise their collective welfare by forming a trade bloc? If the answer is “yes” then a piecemeal enlarge of the bloc will raise bloc members' wellbeing monotonically. Bloc members attain the highest welfare when all nations are part of the bloc. In this world, the formation of a single bloc should trigger a domino effect that leads to worldwide free trade.

The Kemp-Wan theorem tells us that the answer to the question above is always “yes” in a Walrasian setting when nations have access to international lump-sum transfers (Kemp-Wan) or a complete set of commodity taxes and subsidies (Dixit-Norman). Kemp and Wan (1976) – which is probably the first formal contribution to the building/stumbling bloc discussion – makes exactly this point. See Aghion, Antras and Helpman for an elaboration of the Kemp-Wan argument that uses modern cooperative game theory concepts.³¹

While this Kemp-Wan building bloc logic is flawless, it falls down on the very real world problem that nations do not have access to massive lump-sum transfers. Indeed, assuming that such international

³¹ The key concepts are “coalition externalities” (Haberler's spillover) and “grand-coalition superadditivity” (global free trade is first best). The authors assume one nation is the undisputed agenda setter and that unlimited international transfers (transferable utility in game theory parlance) are possible.

transfers are a realistic possibility basically assumes away most the core difficulties facing the international trade system (and international relations more broadly). Without international transfers, the logic of preference-erosion stumbling blocs and cherry-picking stumbling blocs suggests that in many blocs, some bloc members would eventually veto some enlargements.

3.3.3. Veto-avoidance building bloc logic

The exploitation/preference-erosion stumbling bloc logic turned on the fact that bloc members could veto the move to global free trade. The veto-avoidance building bloc logic points out that although the bloc members could veto a multilateral trade liberalisation, they cannot veto further RTAs that may eventually eliminate all tariffs globally. The explosion in the number RTAs among small nations that has been witnessed in the new century may very well be due to a combination of the juggernaut effect and the veto-avoidance logic.

We start by considering the development of an extremely common form of regionalism, namely hub-and-spoke FTAs – where one nation (e.g., US, EU or India) has a network of ‘radial’ bilateral FTAs with some of its trading partners while these trading partners do not have FTAs with each other. Simplicity dictates our continued use of the symmetric Figure 2 framework, so we arbitrarily bestow hub-status on Home.

Incentives for hub-and-spoke FTAs

Roughly, speaking, Home found the bilateral FTA with Partner attractive since the improved market access for Home exporters in Partner’s market more than outweighed the potential welfare losses from trade diversion in Home import market. This suggests that Home might also find a second bilateral FTA with RoW to be welfare enhancing, especially given all the separability that rules Meades secondary and tertiary effects.

As it turns out, in the RTA-diagram framework, Home always gains from signing a second FTA with RoW. Intuitively the point is that Home gains the same preferential market access as it did from the first FTA and it undoes the potentially harmful trade diversion by fully liberalising its import market. To see this in more detail, we re-interpret Figure 6. On the export side, Home’s second FTA wins it preferential access to RoW’s market without giving up its preferences in Partner; this has a net welfare value of areas $+D'+C_1'+C_2'$ in the left panel.³² On the import side, the second FTA brings the price in Home’s market for good 1 to the global free trade level, P^{FT} . The welfare impact of this is the positive trade volume effect area A' plus the conflicting terms of trade effects, areas $-B'$ and $+C_1'$. The formulas in Box 5 show that the overall welfare change for Home is always positive.

Would RoW accept Home’s offer of a second FTA? As it turns out, RoW gains from such an FTA as long as T is not too high (see Box 5 for details). That is, the hub-and-spoke situation is better for RoW than the initially disadvantaged position when it was excluded from the Home-Partner FTA. On the export side, an FTA with Home would improve RoW’s market access a great deal (its export price would rise from the depressed level of $P'-T$ up to the free trade price P^{FT}) and the liberalisation on the import side would have the usual positive trade volume effect and conflicting terms-of-trade effects (identical to those experienced by Home in its first FTA). The threshold tariff for RoW to gain from the hub-and-spoke arrangement is calculated in Box 5.

Plainly Partner will be harmed by the formation of the hub-and-spoke system around Home. Its preferences in Home are eroded and it receives nothing in compensation. Partner would thus like to veto Home’s second FTA, but except in extraordinary circumstances, third nations cannot veto FTAs;

³² Given the separability of the markets the second FTA with RoW would yield a price for good 3 equal to P' .

the main exception is that of customs unions. Functioning customs unions, however, are quite rare in the modern world. Customs unions require supranational decision making capacity to keep all external tariffs in line despite changes in anti-dumping duties, special unilateral preferences to third nations (GSP, etc.) and tariff changes in multilateral trade talks. In fact, the groups of nations that manage such coordination fall into exactly two types: the EU and nations involved in super-hegemon relations (France and Monaco, Switzerland and Liechtenstein, and the South African Customs Union, etc). In fact, the real world is covered with hub-and-spoke trade arrangement, so we assume henceforth that Partner has no veto over Home's FTA policy and the hub-and-spoke system gets set up.

Incentives for spoke-spoke FTAs

The story, however, is not finished. As it turns out, the two spokes may find a spoke-spoke FTA to be advantageous and this would achieve in global duty-free trade (although trade would not necessarily be free due to the exclusion of various 'sensitive sectors', rules of origin and cumulation).

To see this, note that the hub-and-spoke FTA puts Home in an enviable position – giving it the benefits of free trade as far as its imports go and preferential market access for all of its exporters. In this sense, hub-and-spoke bilateralism might be thought of as another example of the exploitation/preference-erosion stumbling bloc logic; Home would veto WTO talks aimed at achieved global free trade. However, this simple world can reach global duty free trade without multilateral talks. An FTA between Partner and RoW would do the job. So would Partner and RoW be interested in an FTA?

Taking hub-and-spoke bilateralism as the point of departure, it is clear that the spokes – Partner and RoW in this example at hand – have a very different view of global free trade than does the hub. Taking Partner as an example, a move to global free trade would do nothing to erode Partner's preferences in Home, since those were already eroded by Home's second FTA. For Partner, the shift to the global-free-trade regime would involve a standard exchange of market access with RoW; Partner would see its export price to RoW rise from P^T to P^{FT} for good 3, and RoW would see a symmetric border price rise for its exports of good 2 to Partner (see Figure 2). The attendant liberalisation of the two nation's import markets would have the usual conflicting trade volume and terms of trade effects, but overall the two nations could find the exchange to be welfare enhancing. In fact, Partner and RoW would always prefer global free trade to the hub-and-spoke situation, as the formulas in Box 5 confirm.

This is certainly not to be taken as a general result. It does, however, illustrate how regionalism could be a building bloc in a world where overall free trade would be in the interest of all nations, but achieving the goal is blocked by nations fearing erosion of their preferences. See Lloyd (2002) for a clear development of the veto-avoidance logic. This line of thinking is one strand in the widely discussed 'competitive liberalisation' logic of Bergsten (1996).

3.3.4. Related logics: induced liberalisation and protection

Before ending this review of the helps-or-hinders literature, it is useful to cover two economic mechanisms that link RTAs and MFN tariffs without formally making the connection with multilateral trade talks. Both mechanisms consider the impact of RTAs on nation's MFN stance in the absence of a new MTN. The first links RTAs to unilateral MFN liberalisation. The second looks at how a RTA can lower or raise a nation's effective MFN tariff rate.

RTAs and unilateral liberalisation: are RTA and MFN tariffs complements or substitutes?

The building-bloc logics examined above directly address the issue of whether RTAs help or hinder the attainment of global free trade. Here we look at a related, but logically distinct question: What is the impact of a RTA on the tariffs a nation would find unilaterally optimal to impose on third nations? Intuitively, the question is whether preferential tariffs are complements or substitutes for MFN tariffs.

The easiest way to organise the various mechanisms is to start from Meade's formula for the welfare impact of any trade policy change in a Walrasian economy, namely T_{od} minus $M_{od}p^*$. A nation choosing its bilateral tariffs optimally would view this as a first order condition and set it to zero to find its optimal tariff. The optimal bilateral tariffs are:³³

$$(2) \quad T_{od} = M_{od} \left(\frac{dp^*}{dM} \right)_{od}$$

where 'o' indicates the origin nation and 'd' the destination nation (i.e. the nation choosing the tariffs). In general, anything can happen to T_{od} when the nation signs a free trade agreement since – according to the Slutsky equation – the direct and cross-good income and substitution effects of the FTA-induced price changes could raise or lower the right-hand side of (2). Attempts to resolve the inherent ambiguity have led to several economic mechanisms being stressed in the literature.

Preferential and MFN tariffs as complements. If the RTA-induced price changes have little impact on the equilibrium slopes of the third-nation import supply curves, then T_{od} is likely to fall since RTAs typically reduce RTA member's trade with third nations (Harberler's spillover), i.e. M_{od} is likely to fall. For example, in the simple RTA-diagram model presented in 2.1, the import supply curves are linear so dp^*/dM does not change, but Harberler's spillover lowers third-nation trade and so Home's optimal tariff on RoW exports fall. Another mechanism that yields complementarity turns on the general principle that taxes become more distortionary when the cross-product variance of rates increases (the so-called uniform tax rate principle). This is a feature of many economic models, especially when administrative and enforcement considerations are taken into account. (This is why most nations impose fairly even indirect tax rates across products.) Since the RTA automatically makes the import tax structure more uneven, there is some presumption that the RTA makes the third nation tariffs more distortionary. In models where this is true, nations are likely to lower third-nation tariffs when it re-optimises its trade tax structure, i.e. RTAs encourage nations to lower applied MFN tariffs.

Preferential and MFN tariffs as substitutes. The most obvious mechanism that suggests the substitutes result (i.e. nations find it optimal to raise third-nation tariffs after having signed a RTA) concerns the market power of the new bloc. If the RTA allows RTA members to better coordinate their third-country tariffs, they are likely to raise external tariffs since they will have more purchasing power than before. This, of course, presumes three fairly unrealistic things: they can coordinate external tariffs, the governments share sufficiently similar objective functions, and their external tariffs are not subject to WTO bindings (or they are willing to violate their WTO commitments). Since most effective RTAs are among developed nations whose tariffs are almost universally bound at near-zero levels (apart from a few low-volume items) and such nations rarely violate their WTO bindings, this mechanism is probably of little real world relevance except in a few commodities (agriculture before the Uruguay Round), and a few low-trade-volume RTAs among developing nations.

Who did what when? Contributions to the literature that have looked at the complementarity-versus-substitutes effects include, inter alia, Reizman (1985), Kennan and Reizman (1990), Krugman (1991, 1993), Bond and Syropoulos (1996a), and Ornelas (2005, 2008).

Imported MFN liberalisation and protection

A closely related line of reasoning considers the automatic impact of FTAs on the external protection of FTA members, when members impose different tariffs on third nations. Under some circumstances, the FTA effectively lowers the higher MFN tariff (imported MFN liberalisation); in other circumstances, the FTA effectively raises the lower MFN tariff (imported MFN protection).

³³ Divide both sides by p_{od}^* to get the optimal ad valorem tariff in terms of the import supply elasticity.

A good example of imported MFN liberalisation can be found in North-South FTAs. The concept can be explained intuitively with reference to Mexico's experience. Mexico signed FTAs with the US and Canada in 1994 (NAFTA) which phased in tariff cuts over 10 to 15 years. Mexican MFN tariffs were (and still are) much higher than US and Canadian MFN tariffs, but as NAFTA brought Mexican prices down to the US internal level, domestic prices in Mexico came to resemble those that Mexico would have observed if it had lowered its MFN tariffs to US levels. To put it differently, the high Mexican MFN tariffs became irrelevant since the same goods could be purchased from the US duty-free and the US internal price was linked to the world price via its low MFN tariff (leaving aside the small sectors still protected by high US MFN tariffs, such as clothing, textiles and footwear). In this sense, Mexico ended up 'importing' the US's low MFN tariffs.³⁴

The argument can be made more precise with Figure 10. The two left panels show import demand of the US (leftmost) and Mexico (middle); the right panel shows the world market for the good under consideration. The US total supply curve is shown in Mexico's panel for reasons that will become clear. The US initially imposes a zero tariff on imports from the rest of the world while Mexico imposes a tariff of T_{MX} on imports from both the US and RoW. When Mexico eliminates duties on US imports, US-made goods can enter Mexico duty-free. Since the Mexican internal price is initially above the US internal price, US firms sell to the Mexican market and in doing so drive down Mexico's internal price to the US internal price – which of course is just the world price. US production entirely displaces Mexican imports from the rest of the world and the Mexican MFN tariff becomes irrelevant.

The feasibility of this outcome is established by noting that the US supply at P^0 is more than sufficient to cover the entire Mexican import demand (point 2 is to the right of point 1). Note that there would be a secondary effect on world prices as the US expands its imports. In the diagram the new world price would be at the intersection of the dotted MD curve and the XS_{RoW} curve. For simplicity's sake, this second order impact is not shown in the two leftmost panels.

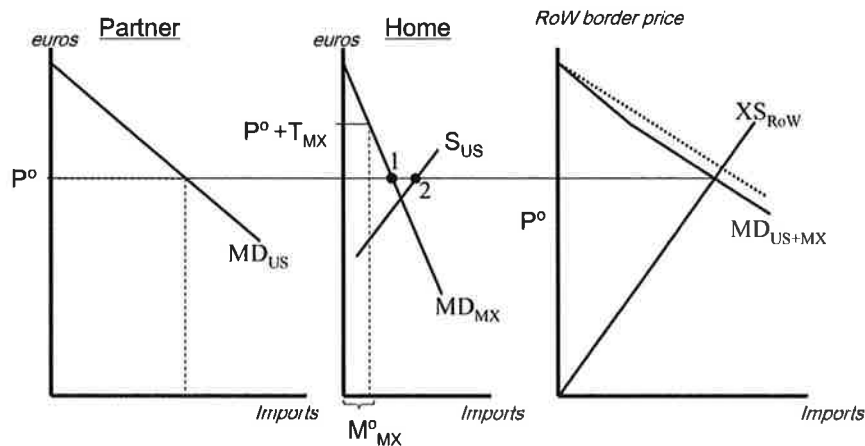


Figure 10: Imported MFN liberalisation

³⁴ Extension of this analysis led to the 'unsustainability of the FTAs' proposition. Vousden (1990 p. 234) argues that Home would be tempted to lower its MFN tariff to just under that of Partner in order to recapture the tariff revenue and Partner would have an incentive to reply, with the resulting race-to-the-bottom tariff making FTAs 'unsustainable'. Richardson (1995) extended and popularised the result. The main results in Shibata (1967), Vousden (1990) and Richardson (1995) – the irrelevance of rules of origin and unsustainability of FTAs – are of little relevance to real-world policy concerns (rules of origin are at the heart of many current policy debates and FTAs, not customs unions, are by far the most prevalent form of PTAs).

If one combines this imported MFN liberalisation with the juggernaut logic, the FTA can eliminate all the firms in Mexico that would have otherwise opposed MFN liberalisation. That is, Mexican industry has no interest in lobbying for the maintenance of high Mexican MFN tariffs since those tariffs provide no protection to Mexican industry. In the case at hand, the Mexican government signed a vast array of FTAs to exchange its now-politically-unless MFN tariffs against preferential access for its exporters.

Since developed nations (the North) tend to have much lower MFN tariffs on most manufactured goods than developing nations (the South), the mechanism suggests that an important implication of North-South FTAs for the world trading system is the way they lower the Southern nation's resistance to further liberalisation. Given that most of the South does not participate in MTN tariff-cutting exercise on the basis of reciprocity, the North-South FTAs are one of the few ways of triggering juggernaut effects in developing nations.

Rules of origin and imported MFN protection. The opposite result – a RTA importing MFN protection to a nation with low MFN tariffs – can occur when highly restrictive rules of origin are imposed. The argument can be illustrated with reference to NAFTA. Since the US's first foray into regionalism – the 1965 US-Canada Auto Pact – US and Canadian rules of origin on autos have been highly restrictive. One of Canada's motives in pushing for the tri-lateralisation of the US-Mexico free trade agreement was to extend its restrictive rules of origin to Mexico and thereby avoid the undermining of the Auto Pact. The rules of origin forced Mexican-based car producers to import parts and components from the US or Canada instead of from third nations. As before, NAFTA equalised US and Mexican internal prices, but this meant that the Mexican prices were linked to the world prices via the higher MFN tariffs in the US and Canada. In this way, 'imported MFN protection' occurred. NAFTA with its rules of origin had effects that mimicked a rise in the Mexican MFN tariffs to the US and Canadian levels. The imported-MFN-protection reasoning is most often associated with Krueger (1993), although it has played a role in the literature discussed in footnote 34.

Although the distortionary impact of rules of origin is limited by the level of the MFN tariff, the all-or-nothing feature of rules of origin for final goods can lead to large 'effective rates of protection.' For example, if a \$20,000 NAFTA-origin car pays zero tariff while the same non-NAFTA car would pay 5%, a rule of origin that stipulates that a particular component must be made inside NAFTA could make it economic to pay up to \$1000 more for the local versus imported component. While the distortion in the final good market is limited to 5%, the distortion in the component market can be much larger (this is the traditional effective rate of protection logic).

Who did what when? Shabata (1967), Vousden (1990), Richardson (1993, 1994, 1995), and Grossman and Helpman (1995) are all important contributors to or users of this line of analysis.

3.4. Bargaining-model stumbling/building bloc logic

The stumbling/building bloc mechanisms discussed above resonant strongly with real-world considerations, in my judgement, since they took advantage of the simple institutions features of real-world tariff cutting in RTAs and MTNs (see Section 3.1). This, however, is not how the Big-Think Regionalism started thinking about the issue (Krugman 1991b, 1993).

As Krugman (1993 p.58) puts it: "In this realm of foggy discussion it is natural for economists to grab hold of any analytic tool they can find, even if they are ill-adapted to the work at hand." I do not believe that the literature that followed up this lead by Krugman has contribution to policy insights since it is built on foundations of sand. Nevertheless, a large literature has developed around it and many academic participants in the Big-Think Regionalism debate refer to the literature's insights, so I will review the basic logic here.

When Krugman wondered how regionalism would affect the GATT, the tool he grabbed for was simple bargaining game theory with two nations that are considering setting tariffs cooperatively (under GATT), or non-cooperatively (Nash tariffs). As Figure 11 shows, both nations prefer the cooperating outcome. He notes: "Trade bargaining ... is characterised by a Prisoners' Dilemma. This Dilemma arises in part from a terms of trade effect of conventional optimal tariff analysis, but also (and presumably in practice mostly) from the effect of each country's tariff on the other country's producer interests." Krugman (1993, p. 72). He goes on to invoke all the usual theorems of repeated games to think about the building/stumbling bloc issue and concludes: "Trade liberalisation must be supported by the belief of countries that if they cheat they will lose from the subsequent collapse of the cooperative outcome."³⁵ The crux of his analysis is to examine the impact that an exogenously formed RTA has on the costs and benefits of cheating.

Terms of trade approach in a nutshell. Much of Krugman's reasoning is informal, so it is worth spelling it out explicitly. The whole analysis turns on three equilibrium welfare levels:

- W^{GATT} is the level of a nation's welfare with global cooperation (GATT tariffs),
- W^{Nash} is national welfare under non-cooperative tariffs, and
- W^{cheat} that reflects the nation's welfare when its government 'cheats', i.e. chooses a tariff to maximise its own welfare when the foreign government embraces its GATT tariffs.

There are two logical steps in the approach. The first step consists of the obvious point that tariffs are worse than a zero-sum game from the global perspective so some form of cooperation could be Pareto improving, but nations have an incentive to cheat. Formally, this is a prisoners' dilemma and it arises when $W^{cheat} > W^{GATT} > W^{Nash}$. The second step involves a dynamic game that models when cooperation is sustained. As Krugman notes, cooperation is self-enforcing when the gains from cheating are more than offset by the losses from the (infinite) punishment. Taking δ as the discount factor, the present value of cooperating forever in symbols is $W^{GATT} / (1 - \delta)$. If cooperation is to be sustained, this must exceed the one-period gain from cheating W^{cheat} plus the present value of the infinite sequence of the Nash outcomes that kick next period after the foreigners realise cheating has occurred, $\delta W^{Nash} / (1 - \delta)$. Clearing the $(1 - \delta)$ terms, the condition for self-sustaining global free trade is:

$$W^{GATT} > (1 - \delta)W^{cheat} + \delta W^{Nash}$$

In words, each nation compares the value of welfare under cooperation to a weighted average of the cheating outcome and the Nash outcome. So far this is a trivial re-labelling of an undergraduate lecture on repeated games.

The contribution of this approach comes in considering how an RTA changes the three levels, W^{cheat} , W^{FT} and W^{Nash} . Krugman (1993) asks whether the formation of a trade bloc among nations make them more or less able/or willing to cooperation. His answer is: It can cut either way.

Krugman's core insight – an insight that has been followed up in a dozen articles since – is that RTAs typically reduces the RTA members' trade with the rest of the world and so reduces both cost and benefit of cheating. Since these work in opposite directions, some bargain-approach papers find that RTAs are building blocks (make cooperation more likely) while other find they are stumbling blocs.

³⁵ This is not a new point. It is very clear in the discussion of Johnson (1953), but probably dates much further back. Indeed, the notion that a *quid pro quo* would be mutually advantageous was probably well understood by trade diplomats as far back as Roman times.

Who did what when? This approach came to be known as the terms-of-trade approach after Kyle Bagwell and Bob Staiger and students formally modelled the issues, starting with Bagwell and Staiger (1993a, b). For examples of this sort application to the regionalism questions see Bond and Syropoulos (1996b), Campa and Sorenson (1996), Bond, Syropoulos and Winters (1996), Bagwell and Staiger (1996), Conconi and Perroni (2000), Conconi (2000), Yi (1996), and Ornelas (2005, 2008).

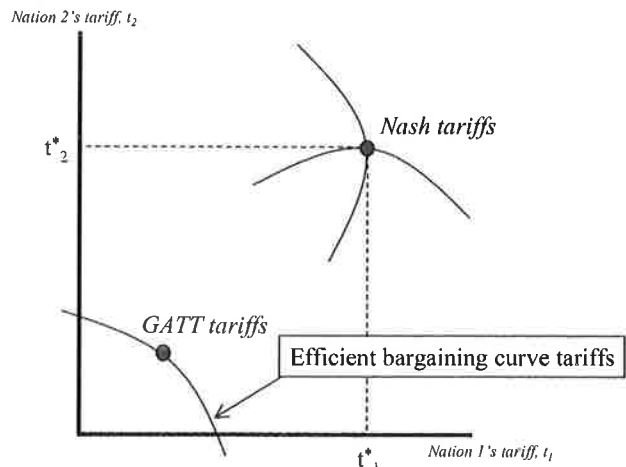


Figure 11: An economic theory of the GATT

Source: Adapted from Krugman (1991b).

3.4.1. Critique of the self-enforcing approach to tariff liberalisation

The approach set out by Krugman seems entirely natural to readers schooled in game theory that was developed to explain strategic interactions among private agents. This sort of model resonates strongly when thinking about how OPEC sustains its cartel, or chemical companies agree to fix the price of vitamin C. When it comes to Big-Think Regionalism, however, it is vitiated by two factually incorrect assumptions: 1) that the cheating period is long enough to make it tempting, and 2) that punishment can be usefully modelled as consisting only of tariff changes.

First, real world tariffs are part of nation's tax code and are thus entirely transparent – at least in the advanced industrialised nations that account for more than 80% of world trade. The domestic laws of the US, EU and Japan, for example, require their governments to publicly announce the tariff schedule. Consequently, tariff cheating is detected by foreigners at exactly the same time the change is implemented (if not before). See Box 1 for a real world example of a surprise tariff cut being detected immediately.

This means that the cheating periods is zero, i.e. δ is unity, so the sustainability condition becomes $W^{GATT} > W^{Nash}$. In other words, as in three stumbling-bloc logics discussed above, nations cooperating if and only if it is individually in their own interest to do so. The whole “self-enforcing” module is irrelevant. This is a lethal problem for the approach since its only contribution was to examine how an RTA affected the costs and benefits of cheating for one period.

The second critical flaw is even more damaging to the relevance of its insights. When Krugman set out the basic approach, he took the repeated Prisoners' Dilemma model straight off the shelf – including its assumption that the punishment can only come in the form of a change in the variable of interest. That is, he assumed that nations can only punish via tariffs. This restriction of the strategy space makes

perfect sense in the analysis of collusion among firms. It does not make sense when it comes to the cooperation in the global trade system.

In the real world, cooperation in the GATT/WTO system is supported by the threat of a breakdown in cooperation far broader than tariffs. The system's members can and do use a wide range of 'carrots and sticks' to induce trade cooperation: foreign development aid, military aid, political support in the international arena, participation in NATO, US troops stationed in Germany and Japan, etc. Moreover if international trade cooperation did breakdown, the costs would far exceed lost markets for exporters. It would halt or seriously hinder international cooperation on many pressing issues – climate change, international terrorism, money laundering, organised crime, humanitarian crisis and control of illegal drugs, to mention just a few. Formally, we would include this in the repeated game by giving the players the ability to impose large costs (unrelated to tariffs) on any player that deviated from cooperation.

In summary, the basic idea that trade cooperation should be modelled explicitly is not useless. Trade cooperation is almost surely dependent upon nations' perceptions of costs and benefits of deviating. Unfortunately, the terms-of-trade approach is focusing on the trees and missing the forest. Its insights as to whether an RTA makes cheating on a tariff more or less attractive misses the point altogether since the main carrots and sticks supporting cooperation have nothing to do with tariffs. This means that RTAs do not have a first order impact on the underpinnings of international trade cooperation – at least not via its impact on the cost and benefit of deviating from cooperation.

The mainstream stumbling/building bloc ideas discussed in Sections 3.2 and 3.3 ignore the sustainability issue. They just assumed that nations would keep their word if they did agree to go to free trade. This strikes me as a fairly reasonable assumption even though threats are explicitly or implicitly part of all forms of cooperation. There is little to gain from explicitly modelling the threats since the most important forms of retaliation are entirely outside the model.

3.5. Other links from RTAs to MTNs

There are a number of points made in the literature that do not fit neatly into the stumbling/building block framework as I have delineated it. One line of reasoning views nation's MFN tariffs, or yes-no stance on multilateral cooperation, as depending upon the strength of various domestic special interest groups. The Big-Think Regionalism question here is whether an RTA weakens or strengthens the pro-trade or anti-trade interest groups. At one end, Winters (1993) argues that regionalism (especially the EU on agriculture) strengthened the hand of protectionists since it worsened Olsen's Asymmetry (i.e. winners from protection are few in number and easy to organise, while the losers are disperse, numerous and difficult to organise politically). He calls this the 'restaurant bill' problem. Just as diners at a table where the bill will be split equally tend to order too much, the EU tended to grant too much protection to farmers. At that other end, De Melo et al (1993), Richardson (1994) and Panagariya and Findlay (1996) argue that an RTA tends to dilute the influence of special interest groups via various mechanisms.

Another important line of thinking asserts that the formation of RTAs creates forces that induce nations to begin and/or finish multilateral trade talks. For example, Lawrence (1991), Sapir (1993), WTO (1995) all argue that the threat of regionalism was a critical element in inducing GATT members to initiate the Uruguay Round and to accept the final Uruguay Round agreement. Windham (1986) makes the same argument for the Tokyo Round. Bergsten (1996) dubs this 'competitive liberalisation' and asserts that regionalism fosters multilateralism and vice versa.

A somewhat related line of thinking, which has not been formalised, is that RTAs are a testing grounds for the GATT/WTO (Lawrence 1996, Bergsten 1996). The prime example here is the EU which dealt with deeper-than-tariff-cutting liberalisation for decades before the issues arrived on the GATT agenda in the Tokyo and Uruguay Rounds. See Ludema (1996) for a partial formalisation of the idea.

Another line of thinking suggests that RTAs can provide commitments that boost the credibility of a nation's policy reforms (Fernandez and Portes 1998). This was explicitly mentioned by Mexico in its requires to the US for an FTA.

4. MULTILATERALISM AND REGIONALISM

The analytic contributions discussed above help frame our thinking on whether regionalism is like to help or hinder multilateral trade liberalisation. The bulk of the theoretical literature over the past two decades, however, has been concerned with an analytically distinct question: Would more regionalism be good or bad for world welfare?

The question of whether RTAs raise or lower welfare is empirical, so one might wonder why so many theoretical articles were published on the topic. The answer lies history-of-thought. When leading thinkers such as Larry Summers, Paul Krugman and Jagdish Bhagwati set the intellectual 'terms of reference' for the regionalism literature in the early 1990s, the EU was the only effective regional arrangement. Since it was widely viewed as unrepresentative, the analysis was conducted by speculating on two scenarios – what the world would look like with and without more regionalism. The comparison of these two speculated scenarios was taken as providing insight on whether spreading regionalism would raise or lower world welfare.

From the perspective of the new century, this literature looks distinctly dated. The world has seen a great deal of regionalism develop and it does not look anything like speculated scenarios of the theoretic models from the 1990s. Since my goal is to review the main theoretical literature, even that which no longer very useful, I cover is-bilateralism-bad papers extracting the literature's main insights without focusing too much on who-did-what-when issues.

4.1. Is Bilateralism Bad?

This literature was kicked off by Paul Krugman's 1991 article entitle "Is bilateralism bad?" (Krugman 1991a). In typical Krugman style, it laid out a seductively simple model, asked it a seductively simple question, and found an answer that was provocative and seductively simple – world welfare would be lowest with three symmetric blocs.³⁶ This provocation produced an impressive string of contributions from authors who felt that Krugman's speculated scenarios model left out critical elements of reality.

To understand the basic logic behind the 3-is-worst result, it is important to note that the simplicity of Krugman's model rules out most of the standard effects. On the production side, he assumes each nation produces a single good, so no production distortions are possible. Since nations and blocs are all symmetric, no nation gains on net from taxing imports. While a tariff tends to lower the cost of imports via the usual terms of trade effect, foreign tariffs exactly offset this since they are the same size and applied to the same amount of trade (each bloc's imports equal its exports). In short, all welfare effects are channelled through the consumer distortion.

³⁶ Note that the exercise was anticipated by Riezman (1985) and Kennan and Riezman (1990), but these did not catch on.

More regionalism with fixed tariffs. To fix ideas, we first hold the tariff constant while varying the size of the symmetric blocs. At one extreme is the case of only 1 bloc in which case there is no tariff and thus no consumer distortion; welfare is maximised. At the other extreme there are zero blocs in the sense that all goods – including domestically produced goods – are taxed at the same rate. Since no production distortion is possible and there is no net terms-of-trade effect, this extreme also attains the first best outcome. After all, consumer distortion is driven by distorted relative prices. Moving away from either extreme lowers welfare. This is the key intuition for the U-shaped relationship between the number of blocs and welfare.³⁷ The fact that the minimum is at 3 is not a result that is robust to parameter changes in Krugman’s own model and it was shown repeatedly to be fall down in less simplified models.

Digging slightly deeper into the intuition, consider what happens when moving from 1 bloc (global free trade) and to 2 and then 3 blocs. When the world forms into two blocs, the price of imported goods rises. Since consumers do not see the true price their nation pays for goods, this leads to socially inefficient expenditure switching; local-good expenditure shares rise and imported-good shares fall. When the world reconfigures into three blocs, two things happen to the consumer distortion. Some goods that were previously untaxed become taxed; this increases the distortion. The distortion between a typical domestic good and a typical imported good, however, falls since more varieties are now taxed. The tension between these two effects is what generates the U-shaped relationship.

Impact of regionalism on tariffs. Krugman assumes the blocs are customs unions and that they set the tariff at the naïve optimal tariff level, i.e. the Nash tariff that is equal to the inverse of the bloc’s import elasticity. Since the bloc gets more market power as it enlarges, Krugman has tariffs rising as the number of blocs declines, although not very much according to his numerical simulations.

4.1.1. Extensions and modifications

A sequence of papers considered more complicated models, with the two main complications being the introduction of national comparative advantages and trade costs.

When authors introduced comparative advantage, it is easy to make the 3-is-worst result go away, but the U-shaped relation remains in all the models that do not allow for production distortions. Bond and Syropoulos (1996a), for example, find that the worst number of blocs may be two, three or more.³⁸

More interestingly, considerations of comparative advantage can reverse Krugman’s result that bigger blocs would like to charge higher tariffs. The intuition can be had by linking the optimal tariff to market power and market power to comparative advantage. When nations produce entirely distinct goods, as in Krugman’s model, grouping nations into, say, two trade bloc does nothing to make the blocs’ production structures more similar and thus market power is merely a matter of size. When

³⁷ Krugman’s deviation is circuitous. The easiest way to see the U-shape is to note that the producer price must adjust so each nation sells its one unit of output. Using the standard CES demand functions, with the price of goods made inside the bloc denoted as ‘ p ’ and the consumer price of exported units equal to $p\tau$, where τ is the tariff factor, i.e. one plus the ad valorem tariff, the market clearing condition for each nation’s one-unit production is

$1 = E(sp^{-\sigma} / P^{1-\sigma} + (1-s)\tau^{-\sigma} p^{-\sigma} / P^{1-\sigma})$ where E is expenditure of a typical nation (normalized to unity), the CES price index is $P^{1-\sigma} = (s + \tau^{1-\sigma}(1-s))p^{1-\sigma}$ and ‘ s ’ is the share of nations inside a typical bloc. Solve for p yields we get $p = (s + (1-s)\tau^{-\sigma}) / (s + (1-s)\tau^{1-\sigma})$. Substituting this into the CES price index, i.e. $P^{1-\sigma} = (s + \tau^{1-\sigma}(1-s))p^{1-\sigma}$, we get $P = (s + (1-s)\tau^{1-\sigma})^{\sigma/(1-\sigma)} (s + (1-s)\tau^{-\sigma})$. Since welfare is inversely proportional to P , we see that welfare is U-shaped in the share of nations that are inside a typical bloc. Note that welfare is unity when s equals zero or unity.

³⁸ Srinivasan (1993) and Deardorff and Stern (1994) provide other examples.

nations' production structures differ in a richer way, formations of blocs can make the blocs' output mix more similar and this will reduce each bloc's market power. For example if there four nations, two labour-poor and two labour-rich, the formation of two blocs with one of each type of nation would produce a world where blocs had equal factor endowments. In a Heckscher-Ohlin world, such a shift would extinguish the market between of the two blocs and with it the incentive to charge a tariff. Of course that is not the only possibility but it shows that the monotonic link between bloc size and optional tariff can be broken. Sinclair and Vines (1995) undertake a more thorough study of the bloc-size-tariff issue, examining, for instance, the case of FTAs as well as customs unions. They make the intuitive point that increasing the size of an FTA does not have a first-order impact on each FTA member's market power and so the tendency for higher tariffs is absent.

The idea that trade costs (i.e. frictional trade barriers) would affect the comparison between regionalism and multilateralism is obvious and was pointed out by Krugman (1991a) in simple extension of his basic model. He assumes that there are continents as well as countries – where continents are defined as groups of nations that have low trade costs among themselves but high trade costs with other continents. If the inter-continental trade costs are high enough, continental free trade comes very close to approximating global free trade (since the remain tariffs fall on very little trade). In the extreme of infinite inter-continental trade cost, the first-best outcome can be obtained with continental trade blocs.

The basic point that 'natural trade blocs' can liberalise most the world's trade had strong resonance with the real-world pattern of RTAs. It also attracted a large number of papers that elaborated on Krugman's basic insight by considering more general sets of trade costs.³⁹

Altogether the Is-Bilateralism-Bad literature is best thought of as a string of parables from which we might glean some insight into real world events. The two main insight, however, are not particular difficult to grasp. The first is that spreading regionalism might or might not harm global welfare. The second is that regional free trade comes closer to mimicking global free trade, the more regionalised is the world trade pattern (due to transport costs or comparative advantage).

5. ENDOGENOUS BLOC FORMATION

In the early 1990s, regional liberalization seemed to be the easy route to integration and trade economists. A central issue in the Big-Think Regionalism literature was therefore why were countries eager to open markets regionally but reluctant to do so multilaterally.

From the new century perspective this question looks odd. In both North America and Europe, regionalism ran into severe political problems (NAFTA, failure to renew fast-track, Maastricht, Nice and Constitutional Treaty rejection, Turkish Enlargement, etc.) while the Uruguay Round passed without much discussion once the final compromises had been completed. For example, NAFTA and the Uruguay Round went into effect in the same year. Moreover, the multilateral trade system continues to enjoy a high-level of background support in all major trading nations. Yet the slowness of the Doha Round, teamed with booming regionalism world wide, has brought discussion of the causes of regionalism back to the forefront of trade policy discussions.

Paul Krugman, Jagdish Bhagwati and other intellectual leaders of Big-Think Regionalism put forth many accounts of the causes of regionalism.⁴⁰ The main explanations were 1) that members were

³⁹ See Frankel, Stein and Wei (1995, 1996), Frankel (1996), Nitsch (1996a, b), Schiff (2001), and Spillembergo and Stein (1995).

frustrated with the GATT's slow progress, especially since the issues had become so much more complex than tariffs and the number of members had grown, 2) that regionalism negotiations were easier, 3) the US conversion from devoted multilateralist to ardent regionalist removed one of the key restraints on regionalism, and 4) the US's aggressive unilateralism scarred Western Hemisphere nations into seeking a 'safe harbour' against the US policies and/or a breakdown in the GATT system.

In a 1993 paper, I proposed a very different explanation with the domino theory of regionalism (Baldwin, 1993, 1995, 1997). The theory views a nation's decision to join an RTA as endogenous and notes that Haberler's spillover provides a de novo political economy force that might make nations change their minds after some of their trade partners form or deepen a preferential trade agreement.

This section reviews the domino theory, its intellectual antecedents and subsequent contributions to the theory of endogenous bloc formation.

5.1. Domino theory

All the contributions in this area require two basic elements – a model of the economy that connects tariff choices to economic outcomes and a political economy model that connects economic fundamental to policy choices. The economic model is not very important as long as Haberler's spillover arises, i.e. as long as the preferential policies of a nation's partners can create export discrimination. The political economy model only requires that exporters tend to support membership in an RTA and import-competitors tend to oppose it. The logic proceeds in two steps. First is the immediate impact of an idiosyncratic deepening of integration in an RTA.

Given an initial political equilibrium where the nation in question has chosen to be outside the RTA, the idiosyncratic deepening or widening of the RTA generates new political economy forces. Specifically, non-member exporters now have a greater stake in membership – they face more discrimination if their nation stays out and greater market access if it joins. Anti-membership forces may also be strengthened in non-member nations, but if the industrial output of export sectors is systematically larger than the output of import-competing sectors (as is usually the case since the export sector produces for both domestic and foreign consumers) and sectors' political power is linked to their size, the shock raises the pro-membership forces more than the anti-membership forces. For outsiders that were previously close to indifferent to membership (politically), these changes shift the domestic political economy equilibrium to the pro-joiners camp.

The second stage starts, if one non-member actually does decide to join. The PTA enlargement implies that discrimination facing the remaining non-members expands and this again heightens the pro-membership political economy forces in outsiders, potentially producing a membership application from an outsider that previously found it politically optimal to stay out. The cycle repeats itself until a new political equilibrium membership in the PTA obtains.

If the world was marked by perfect information and synchronized periodicity in political decision-making, PTA membership bids would be perfectly coordinated and bloc enlargement would happen in a step-like fashion. Uncertainty, imperfect information and mis-matches of decision timings suggest that the new political economy equilibrium may be reached only gradually. During the transition it might look like regionalism was spreading like wildfire.

⁴⁰ See Krugman (1991a, b, 1993), Bhagwati (1991, 1993), Anderson and Blackhurst (1993), Whalley (1996), Lawrence (1996), Bergsten (1996), Panagariya (1996), inter alia.

5.2. The supply side

The domino theory ignores the ‘supply side’ membership – i.e. thinking about whether the incumbents would allow the applicants to join. This was not omission but rather a strategic choice reflecting a judgement that most ‘applications’ to join or form RTAs were driven by economics for the demandeurs (often small nations) and by politics by for the répondeurs. The issue of the economic impact on the répondeurs was thus ignored for parsimony’s sake. Given the rather promiscuous approach that nations are taking in the new century to RTAs, the assumption seems to me as continuing to be relevant.

Much of the subsequent literature on endogenising the membership of regional trade arrangements has focused on putting the supply side into a model with domino-like features. These studies follow the lead of Riezman (1985) in using cooperative game theory to model this “club formation” issue.⁴¹

Yi (1996) was one of the first to formalise the domino logic using cooperative game theory with and without considering the supply constraint. In his model, the domino effect leads to global free trade, if membership is open (i.e. the supply side is ignored), but not if the joiners require the assent of the incumbents and the incumbents care only about how tariffs affect their national welfare. Haveman (1992), Syropoulos (1999), Abrego, Riezman and Whalley (2003), and Melatos and Woodland (2007) consider a number of extension including asymmetric nations.

Aghion, Antras and Helpman (2004) is perhaps the most complete study of what happens when the supply side is fully incorporated. To make headway and reduce proliferation of cases, the allows for transferable utility among nations and assume there is a single nation proposing various configurations. While these assumptions are rather strong and rather at odds with the main outlines of international trade relations in today’s world, they do allow the authors to engage the powerful tools of game theory that was developed for simpler problems. Their main conclusion is that almost anything could happen, but one possible is that – as Yi (1996) found – the domino effect could lead to every nation joining a single RTA.

Asymmetric lobbying. The political economy forces driving the domino effect are strengthened by the peculiar tendency of special interest groups to fight harder to avoid losses than to secure gains. Joining allows excluded firms to avoid damages as well as to win new commercial opportunities, so trade diversion may play a particularly important role in generating new, pro-membership political economy activity. Many explanations for this ‘loser’s paradox’ are possible, but one simple economic interpretation that is relevant to the domino theory is based on unrecoverable investments, i.e. sunk costs. Entry into most industries and markets involves large unrecoverable investments in product development, training, brand name advertisement and production capacity. In such situations, established firms can earn positive profits without attracting new firms, but only in so far as these profits constitute a fair return on the entry investments, i.e. sunk costs create quasi-rents, not pure rents. Given that firms in an industry will have already incurred the sunk costs, deepening of an existing bloc, or formation of a new one will destroy quasi-rents, and thus generate strong, *de novo* political forces pushing the government to redress the new discrimination. The most direct way would be to join the bloc, but other modalities are possible. Government of excluded nations may seek to restore quasi-rents by calling for a multilateral trade round, or forming a new trade bloc among excluded nations.⁴²

⁴¹ Also see Kennan and Riezman (1990), Riezman (1999) and Kose and Riezman (2000).

⁴² See Baldwin and Robert-Nicoud (2002, 2005) for a formal treatment of the ideas and Baldwin (1993b) for an early formal model.

6. DIRECTIONS FOR FUTURE RESEARCH AND CONCLUDING REMARKS

In 2008, the Big-Think Regionalism review might seem like an intellectual indulgence from two distinct angles.

1. Regionalism is here to stay. Regionalism has been raging for two decades and it shows no signs of abating. Since the existing RTAs are not going to be disbanded and more will surely be signed, discussion of whether these RTAs help or hinder multilateral liberalisation is something of a luxury. The real-world issue facing policy makers is how to increase the likelihood that these RTAs help the world trade system and reduce the likelihood that they hurt it. While there may be some room for highly abstract reasoning on this issue, the devil is in the detail. Thinking up ways of making regionalism fit in better with multilateralism will require highly detailed knowledge of matters such as rules of origin, and RTAs' treatment of non-tariff barriers. The chapters in this book provide the most comprehensive effort to gather such information and should therefore provide an excellent springboard for future empirical and theoretical work.

2. Deep, multilateral integration. For the GATT's first 30 years, multilateralism meant shallow integration (mainly tariff cutting). Deep integration (liberalisation of tariffs and many behind-the-border measures) was the purview of regionalism.⁴³ This began to change with the 1979 Tokyo Round although it was not until the Uruguay Round's Single Undertaking that deep integration became an integral part of multilateralism. In 2008, the WTO is dealing with, or talking about dealing with a wide range of areas where liberalisation requires discipline of behind-the-border measures – trade in services, trade-related intellectual property rights, trade-related investment measures, multilateral investment disciplines more generally, technical barriers to trade, government procurement, competition policy, and trade facilitation. A critical issue now facing policy makers is whether the deep integration initiatives in RTAs are throwing up impediments to deep multilateral liberalisation. From this perspective, the exclusive focus on border measures in the help-or-hinder literature seems somewhat amiss.

But does it matter that the stumbling/building bloc theory deals almost exclusively with tariff liberalisation? Cannot one model deep-integration as the removal of barriers that – like tariffs – drive a wedge between internal and external prices? Answering these questions requires detailed knowledge of exactly what sort of deep integration is going on in the RTAs. The chapters in this book provide an excellent starting point. Importantly, the findings suggest that the deep integration the RTAs may indeed be creating incompatibilities that will impede deep integration at the multilateral level – at least in some areas. Several of the chapters provide evidence that at least two 'families' of RTAs emerging as far as deep integration schemes are concerned – one based on the NAFTA model and another based on the EU model. For example, the chapter on technical barriers to trade (TBTs) suggests that one the EU family is fostering a convergence to the EU's product standards and regulations, while the NAFTA family is fostering convergence to international norms. Although the EU and international norms are not always in conflict, it is easy to think of situations where this trend in RTAs will hinder rather than help multilateral TBT liberalisation.

⁴³ Until the 1980s, regionalism principally referred to European PTAs that were implemented (EFTA and the EEC) and a handful of developing nation schemes that were never effectively implemented, the main exception being the ties between Australia and New Zealand. The EEC involved deep integration from its inception, but the trend accelerated with the 1986 Single European Act and its extension to EFTA via the European Economic Area (EEA) negotiations. The 'Down Under' pair embraced deep integration with their 1984 Closer Economic Relationship (CER).

6.1. Future research topics

One of the astounding contrasts that emerge from the detailed mapping of RTAs is the extent to which developing nations accept disciplines in RTAs that they resist at the WTO level. In areas across the board – intellectual property rights, tariff bindings, TBTs liberalisation, access commitments in services, investment agreements, government procurement, subsidy disciplines, trade facilitation, and competition policy – the RTAs reviewed in the chapters show that developing nations have accepted things that they refuse to even discuss at the multilateral level. Does this indicate that something is wrong with the WTO's negotiating procedures? An important topic for future theoretical and empirical research is the identification of the determinants of deep integration in RTAs and the failure of such initiatives at the multilateral level.

A second topic of research concerns the deep difference between the political economy of shallow and deep integration. For example, while discrimination is relatively simple when it comes to tariffs, some behind-the-border measures do not lend themselves to preferences. For example, many TBTs and service trade restrictions are justified on 'good governance' grounds – protection of consumers, etc. – that do not logically admit preferences. If the restrictions are necessary to protect consumers from fraud from one nation, then it seems natural that they should be applied to imports from all partners. Moreover, many behind-the-border measures, especially in services, act to shield incumbents from competition from other domestic firms as well as foreign firms. In such a situation, any liberalisation with respect to foreigners is likely to bring on competition from domestic rivals as well. Such considerations suggest that the trade political economy models of tariff liberalisation need to be modified when thinking about behind-the-border measures.

Theoretical advances in the political economy of various forms of deep integration should be greatly assisted by the mappings contained in the chapters in this book. The near future should see many empirical studies trying to determine why some RTAs include various deep integration initiatives while others exclude them. The answers will surely involve some general considerations – for example, rich nations inevitably place greater trust in the legal and governance structures of other rich nations and such trust is necessary for many of the deeper forms of integration. Other answers, however, are likely to be highly specific to particular areas. For example, services trade liberalisation that goes beyond GATS may be systematically easier between nations that share similar legal and/or educational environments (as is often the case between former colonising nations and their former colonies).

A third topic – one that would involve a high level of abstraction – would be to think about the design of a WTO negotiating framework that would result in successful deep integration in the same way that the GATT's framework fostered successful tariff cutting.

6.2. Concluding remarks

The bulk of this paper presented a survey regarding the analytics of the "classic" debate on regionalism versus multilateralism. The intent of the paper, however, was to suggest that there is a need to move the literature's focus from the high theory of shallow integration to a more policy-relevant issue – the theory and empirics of deep integration in regional versus multilateral contexts. There is also a need to advance the profession's thinking on how the liberalisation in RTAs can be made to be more supportive of multilateral liberalisation, i.e. on how one can promote convergence/harmonization of RTAs. The vast datasets that are contained in the other chapters in this book provide a rich stimulus to future research that goes beyond effect.

Box 1: Nixon's tariff 'cheating'

A quite extreme example can be found in the surprise announcement of a 10% rise in all US tariffs on 15 August 1971. The announcement was made on national television by President Nixon, acting under that the emergency authority in the Trade Expansion Act (authority that was inserted to allow the US to instantly respond to changes in the international environment, including sudden tariff hikes by foreigners). The nightly news coverage of the announcement included interviews with European and Japanese policy makers discussing the policy change even before it was implemented the next morning. The only other legal option the US had was an act of Congress, which would have been equally public.

Interestingly, the 10% surcharge was to take immediate effect, but bureaucratic delays in US Customs and confusions over the exact coverage meant that implementation was not immediate. The reaction by foreigners, however, was immediate and the 'import surcharge' was formally removed after just four months. US tariff revenue as a percent of US imports was lower in 1971 than in 1970.⁴⁴

While this episode demonstrates that nations have and might use tariffs as a stick to get their way (in this case, the problems was the European and Japanese refusal to revalue their currencies despite the growing US trade deficit), it also clearly illustrates that there is no period in which the cheater could profit before the others work out what has been done.

Box 2: Viner Verbatim

Box 3: Walrasian net welfare effects: linear example

The appendix works through the general derivation of the modern analysis of a marginal tariff liberalisation. This box works through a linear example that makes the reasoning more transparent and directly connects it to the standard undergraduate analysis involving consumer surplus, producer surplus and tariff revenue. It also specifies the translation from marginal tax analysis into discrete tariff changes. We work in a 3 nation Walrasian setting with tariffs as the only distortion. We assume Home imports from both Partner and RoW nations and initially applies the same specific tariff T to imports from both sources. Assuming a linear demand function, $p=1-C$, where C is Home consumption and p is the Home's internal price, and a linear supply function, $p=b+Q$, where Q is Home production, the sum of consumer surplus, producer surplus and tariff revenue is:

$$W=(a-p)/2+(p-b)(Q/2)+(p-p^*)(C-Q)$$

where p^* is the common border price and the three terms are, respectively, consumer surplus, producer surplus and tariff revenue ($C-Q$ equals imports). After the tariff liberalisation, welfare can be written as $W'=(a-p')/2+p'(Q'/2)+(p'-p^*)(C'-Q')$, where the 'primes' indicate post-liberalisation quantities and prices. Subtracting the post-liberalisation welfare, W' , from the pre-liberalisation W and re-arranging:

$$W=(p'-p)(M'-M)/2 - (p^*-p^*)M$$

where $M=C-Q$ and $M'=C'-Q'$. Intuitively, the traditional focus on consumer surplus, producer surplus and tariff revenue is unnecessarily cumbersome since the incidence of the tariff can always be decomposed into a part that falls on domestic agents (consumers) and foreign agents (foreign exporters). That is, the incidence of the tariff on consumers is an internal transfer; only the incidence on foreigners (via change in the border price) matters for net welfare.

⁴⁴ See <http://www.state.gov/r/pa/ho/frus/nixon/iii/8698.htm>. You can watch Nixon announce it at <http://thamessthatgreenspanmade.blogspot.com/2008/03/nixon-ends-gold-convertibility.html>.

Box 4: The Walrasian RTA diagram model in maths

The three goods (1, 2 and 3) are produced according to rising marginal costs. For good 1, we assume Partner and RoW have comparative advantage, while Home does not. The assumed linear supply curves that yield this are:

$$(3) \quad p = b + X_1^H, \quad p = X_1^P, \quad p = X_1^R; \quad 0 < b$$

where the superscript indicates the producing nation and the subscript indicates the good. Demand for good 1 is identical in all nations and given by:

$$(4) \quad p_1^j = a - C_1^j, \quad j = H, P, R; \quad 0 < b < a$$

The supply and demand curves for good 2 and 3 are similar, but Partner has the comparative disadvantage in good 2 and RoW in good 3. The demand curves are identical for all goods.

As is well known, the utility function that generates this demand structure involves quasi-linear preferences (utility is separable and quadratic in consumption of each of the 3 goods). The model has a fourth untaxed good that is only introduced to formally eliminate Meade's tertiary effects (it is produced under constant returns using only labour and identical technology worldwide thus ensuring trade balance as long as parameters are such that all 3 nations produce some in all equilibriums, an assumption we maintain throughout).

We take labour as numeraire and choose units so that one unit of labour is needed to produce one unit of the fourth good. The supply curves are thus the marginal cost curves where marginal cost is, e.g. the wage 'w' times total output X for nations with a comparative advantage in the good in question. As usual, rising supply curves generate producer surplus which we take to be the reward to the implicit, scarce specific factor that is generating the diminishing returns.

As is well known, consumer surplus plus producer surplus plus tariff revenue is an exact measure of utility, assuming tariff revenue is returned lump sum to the workers/owners.

Box 5: The Walrasian RTA diagram model: equilibrium prices and welfare

The model described in Box 5 is a system of linear equations. The equilibrating variable is the internal price in the nation that imports the good (1, 2 or 3; the price of good 4 is always unity). Solving for typical good 1 when $T=0$, the equilibrium price for the typical good and the equilibrium welfare of a typical nation (Home) are:

$$(5) \quad p_1^{FT} = \frac{3a + b}{6}, \quad W_H^{FT} = \frac{9a^2 - 6ab + 5b^2}{12}$$

The equilibrium price is identical for goods 2 and 3; internal and border prices are equal.

When nations impose a symmetric MFN tariff T , we get the MFN equilibrium price, p_1^{MFN} , and when Home and Partner sign a bilateral FTA, we get the FTA equilibrium price p_1^{FTA} . Direct calculation shows these to be:

$$(6) \quad p_1^{MFN} = p_1^{FT} + \frac{2}{3}T; \quad p_1^{FTA} = p_1^{FT} + \frac{1}{3}T$$

The MFN price applies to all three goods, but the FTA price concerns only the liberalised goods (good 1 which Home imports from Partner and good 2 which Partner import from Home). Under the bilateral FTA, the price of good 3 remains at the MFN level (additive separability and free trade in the 4th good

eliminate all cross-good effects). The prohibitive MFN tariff is $T=1/2b$; we assume throughout that $T < 1/2b$. The so-called optimal tariff (Nash tariff) is $T=1/8b$.

The welfare of a typical FTA partner, Home, under MFN tariffs and bilateral FTA are by the first two expressions in:

$$(7) \quad W_H^{MFN} = W_H^{FT} - \frac{2}{3}T, \quad W_H^{FTA} = W_H^{FT} + \frac{b}{9}T - T^2, \quad W_R^{FTA} = W_R^{FT} - \frac{5b}{9}T^2$$

The third expression gives the welfare of the excluded nation under the FTA.

The welfare levels of Home and RoW under the hub-and-spoke FTA arrangement are:

$$(8) \quad W_H^{H\&S} = W_H^{FT} + \frac{2b}{9}T + \frac{2}{9}T^2, \quad W_R^{H\&S} = W_R^{FT} - \frac{b}{9}T - \frac{7}{9}T^2$$

Comparing RoW welfare under free trade and H&S, we see the gain from moving to free trade is $=bT/9 + 7T^2/9 > 0$.

Box 6: Preferential liberalisation in the Brander-Krugman model

One set of stumbling/building bloc papers employ the Brander-Krugman model, i.e. Cournot oligopoly with segmented markets. As is well known, reciprocal dumping trade among all nations arise in this model for a wide range of parameters. This set up is useful since every nation sells to every other nation. Moreover, the Cournot oligopoly and market segmentation immediately give us Smith's certitude and Haberler's spillover. Interestingly, Viner's ambiguity does not arise in the simple symmetric version since the procompetitive effects of preferential liberalisation are always strong enough to outweigh any terms-of-trade losses for the integrating nations. One important drawback in this approach is the inability to distinguish between import-competing firms and export firms (they are the same firms).

On the demand side, the model is quite similar to the Walrasian model in Box 4. It has three nations (Home, Partner and RoW) and uses the same quasi-linear preferences with an untaxed numeraire to neutralise Meade's secondary and tertiary effects. The demand curves are identical to (4). Marginal costs are flat (set to zero without loss of generality, given the linear demands), there is one firm per nation, and firms play Cournot in all markets (assumed to be segmented).

Solving the Nash first order conditions, profits of a Home firm depend upon tariffs according to:

$$(9) \quad \pi_H = \frac{1}{16} \left((a + t_{PH} + t_{RH})^2 + (a + t_{RP} - 3t_{HP})^2 + (a + t_{PR} - 3t_{HR})^2 \right)$$

where the subscripts on the directional specific tariffs (the t 's) are "from, to" (e.g. t_{RP} is the tariff on the good shipped from RoW to Partner). The expressions for Partner and RoW profits are isomorphic. Equilibrium welfare of a typical nation – defined as the sum of consumer and producer surplus plus tariff revenue – depends upon the matrix of tariffs according to:

$$(10) \quad W_H = \frac{1}{32} \left\{ 15a^2 + t_{HP}(18t_{HP} - 12(a + t_{RP})) + t_{HR}(18t_{HR} - 12(a + t_{PR})) + t_{HR}(18t_{HR} - 12(a + t_{PR})) + t_{RP}(4a + 2t_{RP}) + t_{PR}(4a + 2t_{PR}) + t_{RH}(6a - 21t_{RH}) \right\}$$

Krishna (1998) uses profit at the government's objective function, while Freund (2000) uses welfare.

The value of Home profits under the three regimes – symmetric MFN tariffs ‘MFN’, FTA with Partner ‘FTA’, and hub-and-spoke FTA with Partner and RoW ‘HS’– are:

$$(11) \quad \pi_H^{MFN} = \pi_H^{FT} + \frac{3t^2 - at}{4}, \quad \pi_H^{FTA} = \pi_H^{FT} + \frac{3t^2}{8}, \quad \pi_H^{HS} = \pi_H^{FT} + \frac{t^2 + 2at}{8}$$

where $\pi_H^{FT} = \frac{3a^2}{16}$. The corresponding expressions for Partner are the same for the MFN and FTA cases (by symmetry of nations), and RoW’s expression is also identical for the MFN case. The profits for the three remaining cases are:

$$(12) \quad \pi_P^{HS} = \pi_H^{FT} + \frac{5t^2 - 2at}{8}, \quad \pi_R^{FTA} = \pi_H^{FT} + \frac{11t^2 - 4at}{8}, \quad \pi_R^{HS} = \pi_H^{FT} + \frac{5t^2 - 2at}{8}$$

Box 7: Juggernaut effect maths

See “A Simple Model of the Juggernaut Effect of Trade Liberalisation, CEP DP No. 0845

Appendix A: APPENDIX: DERIVATION OF THE WELFARE ORGANISING FRAMEWORK

See the on line version.

REFERENCES

- Aghion, P., P. Antràs and E. Helpman (2004). Negotiating Free Trade. NBER Working Paper no.10721, National Bureau of Economic Research, Cambridge, MA.
- Anderson, K and Blackhurst, R. (1993), 'Introduction and summary', in K.Anderson and R.Blackhurst (eds.), Regional integration and the global trading system, (London: Harvester-Wheatsheaf).
- Bagwell, K. and Staiger, R. W. (1996) "Regionalism and Multilateralism Tariff Cooperation," Paper prepared for International Economics Association Round-Table Conference, International Trade Policy and the Pacific Rim, 15-17 July, pages 37.
- Bagwell, K. and Staiger, R. W. (1996) “Preferential Agreements and the Multilateral Trading System,” mimeo.
- Bagwell, K. and Staiger, R. W. 1993a. “Multilateral Tariff Cooperation during the Formation of Regional Free Trade Areas,” NBER Working Paper Series 4364.
- Baier, S.L., and J.H. Bergstrand (2003). Economic Determinants of Free Trade Agreements. Journal of International Economics, vol. 64, pp. 29-63.
- Balassa, Bela (1967). “Trade Creation and Trade Diversion in the European Common Market.” The Economic Journal, Vol. 77.
- Baldwin, R. (1993), 'Asymmetric Lobbying: Why Governments Pick Losers,' GIIIS mimeo available by email.
- Baldwin, R.E., Haaparanta, P. and Kiander, J. 1995. Expanding Membership in the European Union, Cambridge University Press, Cambridge.

- Baldwin, Richard (1993). "A domino theory of regionalism," CEPR Discussion Paper 857; NBER Working Paper, 4465.
- Baldwin, Richard (1994). Towards an integrated Europe, CEPR (London).
- Baldwin, Richard (1995). "A domino theory of regionalism," Chapter 2 of Baldwin, Haaparanta and Kiander (eds.), pp. 25-48. Republished as Chapter 23 (pp 479-502) in J. Bhagwati, P. Krishna and A. Panagariya (1999), Trading blocs: Alternative approaches to analyzing preferential trade agreements, Cambridge: MIT Press.
- Baldwin, Richard (1997). "The causes of regionalism," The World Economy, Volume 20, Number 7, pp. 865-888(24).
- Baldwin, Richard, and Charles Wyplosz, (2006). The Economics of European Integration, McGraw-Hill, 2nd Edition.
- Baldwin, Robert (1970). Non-Tariff Distortions of International Trade. Washington, D.C.: The Brookings Institution.
- Baldwin, Robert (1985). The Political Economy of U.S. Import Policy. Cambridge, MA: The MIT Press.
- Bergsten, F. (1996), 'Competitive liberalisation and global free trade: A vision for the early 21st century,' (Washington DC: Institute for International Economics).
- Bergsten, Fred (1991). "Commentary: The Move toward Free Trade Zones", Economic Review, Federal Reserve Bank of Kansas City, issue Nov, pages 27-35.
- Bergsten, Fred (1996). "Globalizing Free Trade: The Ascent of Regionalism," Foreign Affairs, May/June.
- Bergstrand, J. and S. Baier "Economic Determinants of Free Trade Agreements," Journal of International Economics, Vol 64, No. 1, October 2004, 29-63.
- Bhagwati, J. (2008). "Termites in the Trading System" Oxford University Press.
- Bhagwati, J. (1993). "Regionalism and multilateralism: an overview," in J. de Melo and A. Panagariya (eds), New Dimensions in Regional Integration, Cambridge University Press, Cambridge.
- Bhagwati, J. N. and Panagariya, A. 1996. "Preferential Trading Areas and Multilateralism: Strangers, Friends of Foes?," in Bhagwati and Panagariya (eds.) Free Trade Areas or Free Trade? The Economics of Preferential Trading Agreements, AEI Press, Washington D.C. (forthcoming).
- Bhagwati, J.N. 1991. The World Trading System at Risk, Harvester Wheatsheaf, Hemel Hempstead.
- Bond, E.W. and Syropoulos, C. 1996b. "Trading Blocs and the Sustainability of Inter-regional Cooperation," forthcoming in Canzoneri, M., Ethier, W. and Grilli V. (eds.) The New Transatlantic Economy, Cambridge University Press, Cambridge.
- Bond, E.W., Syropoulos, C. and Winters, L.A. 1996. "Deepening of Regional Integration and External Trade Relations," CEPR Discussion Paper No. 1317.
- Bond, Eric W., and Constantinos Syropoulos (1996). "The size of trading blocs: Market power and world welfare effects," Journal of International Economics, Vol. 40, pp. 411-437.
- Cadot, O., De Melo, and Olarreaga, J. 1996. "Regional Integration and Lobbying for Tariffs Against Non-Members," mimeo, University of Geneva.
- Campa, J.M.. and Sorenson, T.L. (1996) "Are Trade Blocs Conducive to Free Trade?," The

- Cheng, I., and H. Wall (2005), Controlling for Heterogeneity in Gravity Models of Trade and Integration. Federal Reserve Bank of St. Louis Review, vol. 87, pp. 49-63.
- Collie, D. R. (1997). "Bilateralism is Good: Trade Blocs and Strategic Export Subsidies", Oxford Economic Papers 49, 504-520.
- Conconi, P., 2000. "Trade Bloc Formation Under Imperfect Competition," The Warwick Economics Research Paper Series (TWERPS) 571, University of Warwick, Department of Economics.
- Conconi, P., and C. Perroni (2000). "Issue Linkage and Issue Tie-in in International Negotiations", Fondazione Eni Enrico Mattei, Note di Lavoro: 57/2000.
- Cooper, C. and B. Massell (1965). "A new look at customs union theory," Economic Journal, 75, pp 742-747.
- Croome, John (1995). Reshaping the World Trading System: A History of the Uruguay Round, DIANE Publishing, New York.
- Deardorff, A.W. and Stern, R.M. 1994. "Multilateral Trade Negotiations and Preferential Trading Arrangements," in Deardorff, A.V. and Stern, R.M., Analytical and Negotiating Issues in the Global Trading System, Ann Arbor: University of Michigan Press.
- Destler, I.M. (1986). American Trade Politics, Institute for International Economics, Washington DC.
- Fernandez, Raquel, and Jonathan Portes. 1998. "Returns to Regionalism: An Analysis of Nontraditional Gains from Regional Trade Agreements." World Bank Economic Review 12 (2, May): 197-220.
- Finger, Michael, Ulrich, Reincke and Adriana Castro (1999). "Market access bargaining in the Uruguay Round - Rigid or relaxed reciprocity?" Policy Research Working Paper Series 2258, The World Bank.
- Frankel, J.A. 1996. Regional Trading Blocs, forthcoming, Institute for International Economics, Washington, D.C.
- Frankel, J.A. and Wei, S.J. 1996. "Regionalization of World Trade and Currencies: Economics and Politics," Chapter 7 of Frankel (ed.) The Regionalization of the World Economy, forthcoming, Chicago University Press.
- Frankel, J.A., Stein, E. and Wei, S. J. 1995. "Trading Blocs and the Americas: The Natural, the Unnatural, and the Super-Natural," Journal of Development Economics, 47, 61-95.
- Frankel, J.A., Stein, E. and Wei, S.J. 1996. "Continental Trade Blocs: Are They Natural and Super-Natural?," Chapter 4 of Frankel (ed.) The Regionalism of the World Economy, forthcoming, Chicago University Press.
- Freund, C. (2000a) "Different Paths to Free Trade: The Gains from Regionalism". Quarterly Journal of Economics 115:4, 1317-1341.
- Freund, C. (2000b). "Multilateralism and the endogenous formation of preferential trade agreements", Journal of International Economics 52 (2000) 359-376.
- Gatsios, K. and Karp, L. 1995. "Delegation in a General Equilibrium Model of Customs Unions," European Economic Review, 39(2), pp. 319-33.
- Goto, J. and Hamada K. 1995. "EU, NAFTA and Asian Responses: A Perspective from the Calculus of Participation," NBER Working Paper No. 5325.

- Goto, J. and Hamada, K. 1994. "Economic Integration and the Welfare of those who are left Behind: An Asian Perspective," RIEB Kobe University, Discussion Paper No. 47.
- Grossman, G. and E. Helpman (1995). "The Politics of Free-Trade Agreements," *American Economic Review* 85 (1995b):667-690.
- Grossman, G. and Helpman, E. 1994. "Protection for Sale," *American Economic Review*, 85, 667-90.
- Hathaway, Oona (1998). "Positive Feedback: The Impact of Trade Liberalization on Industry Demands for Protection," *International Organization*, Vol. 52, No. 3. (Summer, 1998), pp. 575-612.
- Helpman, E. 1995. "Politics and Trade Policy," CEPR Discussion Paper No. 1269.
- Holmes, Tammy (2005). *What Drives Regional Trade Agreements that Work?* HEI Working Paper, No. 07/2005, Graduate Institute of International Studies, Geneva.
- Hufbauer, G. (1989). *Background Paper for The Free Trade Debate*, Reports of the 20th Century Fund Task Force on the Future of American Trade Policy, Priority Press, New York.
- Hufbauer, G.C. 1990. *Europe 1992: An American Perspective*, Brookings, Washington, D.C.
- Irwin, D.A. 1993. "Multilateral and Bilateral Trade Policies in the World Trading System: A Historical Perspective," Chapter 4 of De Melo and Panagariya (eds.), pp. 90-119.
- Irwin, Douglas (1993). 'Multilateral and bilateral trade policies in the world trading system: an historical perspective,' in De Melo, J. and Panagariya, A. (eds), *New dimensions in regional integration*, (Cambridge: Cambridge University Press for CEPR).
- Jackson, J.H. 1991. "The European Community and World Trade: The Commercial Policy Dimension," Discussion Paper No. 298, Institute for Public Policy Studies, University of Michigan, Ann Arbor.
- Jacquemin, A. and Sapir, A. 1988. "European Integration or World Integration?," *Weltwirtschaftliches Archiv*, vol. 124, pp. 127-39.
- Johnson, H. (1953). "Optimum tariffs and retaliation," *Review of Economic Studies*, 21, 2, pp 142-53.
- Johnson, H. (1960a). "The economic theory of customs unions." *Pakistan Economic Review*, 10, pp 14-30. Reprinted in Johnson (1962).
- Johnson, H. (1960b). "The cost of protection and the scientific tariff." *Journal of Political Economy*, 68, pp 327-345.
- Johnson, H. (1965). "An economic theory of protectionism, tariff bargaining and the formation of customs unions." *Journal of Political Economy*, 73, pp 256-283.
- Karacaovali, Baybars and Nuno Limão (2005). "The Clash of Liberalizations: Preferential vs. Multilateral Trade Liberalization in the European Union," CEPR Discussion Papers 4973.
- Katz, E. (2001). *Bias in Conditional and Unconditional Fixed Effects Logit Estimation*. *Political Analysis*, vol. 9, pp. 379-384.
- Kemp, M.C. and Wan, H., Jr. 1976. "An Elementary Proposition Concerning the Formation of Customs Union," *Journal of International Economics*, vol. 6, pp. 95-7.
- Kennan, J. and Riezman, R. 1990. "Optimal Tariff Equilibria with Customs Unions," *Canadian Journal of Economics*, 23, 70-83.
- Keohane, R.O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*, Princeton University Press, Princeton, N.J.

- Kowalczyk, C, and Sjoström, T. 1992. "Bringing GATT into the Core," *Economica*, vol. 61, pp. 301-17.
- Kowalczyk, C. (1990). "Welfare and customs unions," NBER Working Paper, 3476.
- Krishna, P. (1998). "Regionalism and Multilateralism: A Political Economy Approach," *Quarterly Journal of Economics*, February 1998, Vol. 113, No. 1, Pages 227-250.
- Krishna, P., and Panagariya, A., 2000, "A Unification of Second-Best Results in International Trade," *Journal of International Economics*.
- Krugman P. (1991a). "Is Bilateralism Bad?," in Helpman and Razin.(eds) *International Trade and Trade Policy*, Cambridge MA, MIT Press.
- Krugman, Paul (1991b). "The move toward free trade zones," *Economic Review*, Federal Reserve Bank of Kansas City, issue Nov, pages 5-25.
- Krugman, Paul (1993), 'Regionalism versus multilateralism: Analytic notes', in De Melo, J. and Panagariya, A. (eds), *New dimensions in regional integration*, (Cambridge: Cambridge University Press for CEPR).
- Lawrence, R. (1996), *Regionalism, multilateralism and deeper integration*, (Washington: Brookings Institute).
- Lawrence, R.Z. 1991. "Emerging Regional Arrangements: Building Blocs or Stumbling Blocks?" in O'Brien (ed.) *Finance and the International Economy 5: The AMEX Bank Review Prize Essays*, New York: Oxford University Press, pp 23-35.
- Levy, P.I. 1994. "A Political Economy Analysis of Free Trade Agreements," *Economic Growth Centre Discussion Paper No. 718*, Yale University.
- Levy, P.I. 1996a. "Lobbying and International Cooperation in Tariff Setting, mimeo, Yale University.
- Levy, P.I. 1996b. Comment on "Regionalization of World Trade and Currencies: Economics and Politics," Chapter 7 of Frankel (ed.) *The Regionalization of the World Economy*, forthcoming Chicago University Press.
- Levy, Philip, 1997, "A political-economic analysis of free-trade agreements," *American Economic Review*, Vol. 87, pp. 506–519.
- Limão, Nuno (2006). "Preferential Trade Agreements as Stumbling Blocks for Multilateral Trade Liberalization: Evidence for the United States," *American Economic Review*, 96, 3, pp 896-914.
- Limão, Nuno, 2006, "Are Preferential Trade Agreements with Non-trade Objectives a Stumbling Block for Multilateral Liberalization?" *Review of Economic Studies*.
- Lloyd, Peter John (2002). "New Bilateralism in the Asia–Pacific" *The World Economy*, Vol. 25, pp. 1279-1296.
- Ludema, R. 1996. "On the Value of Preferential Trade Agreements in Multilateral Negotiations," *Journal of International Economics*.
- Mansfield, E.D. and E. Reinhardt (2003). *Multilateral Determinants of Regionalism: The Effects of GATT/WTO on the Formation of Preferential Trade Agreements*. International Organization, vol. 57, pp. 829-862.
- Maskus, K. and J.S. Wilson (eds) (2001). *Quantifying the impact of technical barriers to trade: Can it be done?*, Michigan University Press, Ann Arbor.

- Mclaren, J. (2002) A Theory of Insidious Regionalism. *Quarterly Journal of Economics* 117:2, 571-608.
- Melatos, Mark & Woodland, Alan, 2007. "Endogenous trade bloc formation in an asymmetric world," *European Economic Review*, Elsevier, vol. 51(4), pages 901-924, May.
- Messerlin, P.A. 1983. "Bureaucracies and the Political Economy of Protection: Reflections of a Continental European," *Weltwirtschaftliches Archiv*, 117, pp. 468-96.
- Nitsch, V. (1996) "Do Three Trade Blocs Minimize World Welfare?," *Review of International Economics*, Vol. 4, pp 355-363.
- Nitsch, V. (1996) "Natural Trading Blocs: A Closer Look," prepared for European Economic Association Meeting, Istanbul, August.
- Nogues, J. and Quintanilla, R. 1993. "Latin America's Integration and the Multilateral Trading System," in De Melo and Panagariya (eds.) *New Dimensions in Regional Integration*, pp. 278-313.
- Nordstrom. H. 1995. "Customs unions, regional trading blocs and welfare," Chapter 3 of Baldwin, Haaparanta and Kiander (eds.), pp. 54-78.
- Oye, K. 1992. *Economic Discrimination and Political Exchange: World Political Economy in the 1930s and 1980s*, Princeton University Press: Princeton, NJ.
- Panagariya (1996), 'APEC and the United States,' CIES DP, (University of Adelaide).
- Panagariya, A. and Findlay, R. 1994. "A Political Economy Analysis of Free Trade Areas and Customs Unions," Policy Research Working Paper 1261, World Bank, Washington D.C.
- Perroni, C. and Whalley, J. 1994. "The New Regionalism: Trade Liberalization or Insurance?," NBER Working Paper No. 4626.
- Richardson, M and Desruelle, D. 1994. "Fortress Europe: Jericho or Chateau d'If?," mimeo.
- Richardson, M. 1993. "Endogenous Protection and Trade Diversion," *Journal of International Economics*, 34(3-4), 309-24.
- Richardson, M. 1994. "Why a Free Trade Area? The Tariff Also Rises," *Economics and Politics*, 6(1), pp 79-96.
- Richardson, M. 1995. "Tariff Revenue Competition in a Free Trade Area," *European Economic Review*, vol. 39, pp. 1429-37.
- Rieder, R. (2006). *Playing Dominoes in Europe: An Empirical Analysis of the Domino Theory for the EU, 1962-2004*. HEI Working Paper, No. 11/2006, Graduate Institute of International Studies, Geneva.
- Riezman, R. (1999) "Can Bilateral Trade Agreements Help Induce Free Trade", *Canadian Journal of Economics*, 32(3), pp. 751-766.
- Riezman, R. 1985. "Customs Unions and the Core," *Journal of International Economics*, 19, 355-65.
- Sapir, A. (1997), 'Domino effects in Western European Trade, 1960-1992,' CEPR DP (London). Published as 'Domino effects in Western European regional trade, 1960-92', *European Journal of Political Economy*, Vol 17: 386, 2001.
- Sapir, A. 1993. "Discussion of Chapter 7," De Melo and Panagariya (eds.), pp. 1491-1506.
- Schattschneider, E.E. 1935. *Politics, Pressures and the Tariff*, New York: Prentice-Hall.

- Serra et al. (1997), *Reflections on regionalism: Report of the study group on international trade*, (New York: Carnegie Endowment for International Peace).
- Shepsle, K.A. and Weingast, B.R. 1981. "Political preferences for the pork barrel: a generalisation," *American Journal of Political Science*, 25, pp. 96-111.
- Sinclair, P. and Vines, D. 1995. "Bigger Trade Blocs Need not Entail more Protection," mimeo, University of Birmingham.
- Spilimbergo, A. and Stein, E. 1996. "The Welfare Implications of Trading Blocs among Countries with Different Endowments," Chapter 5 of Frankel (ed.) *The Regionalization of the World Economy*, forthcoming, Chicago University Press.
- Srinivasan, T.N. 1993. "Regionalism vs. Multilateralism: Analytical Notes' Comment," in De Melo and Panagariya (eds) *New Dimensions of the Regional Integration*, pp 84-89.
- Summers, L. (1991), "Regionalism and the World Trading System." *Economic Review*, Federal Reserve Bank of Kansas City, issue Nov, pages 295-302.
- Viner, J. (1950). *The Customs Union Issue*. Carnegie Endowment for International Peace, New York.
- Whalley, J. (1993), 'Regional trade arrangements in North America: CUSTA and NAFTA,' in De Melo, J. and Panagariya, A. (eds), *New dimensions in regional integration*, (Cambridge: Cambridge University Press for CEPR).
- Whalley, J. (1996), 'Why do countries seek regional trade agreements?' NBER WP 5552 (Cambridge MA)
- Winham, G.R. 1986. *International Trade and the Tokyo Round Negotiation*, Princeton University Press, Princeton, N.J.
- Winters, A. (1996), 'Regionalism versus multilateralism,' World Bank Policy Research Working Paper 1687, (Washington).
- Winters, Alan (1991). *International Economics*, 4th Edition, Routledge, London.
- Winters, L.A. 1993. "The European Community: A case of Successful Integration," CEPR Discussion Paper No. 755.
- Winters, L.A. 1994. "The EC and Protection: The Political Economy," *European Economic Review*, 38, pp 596-603.
- Winters, L.A. 1995. "Who Should Run Eastern European trade policy and How," Chapter 2 of Winters (ed.) *Foundations of an Open Economy: Trade Laws and Institutions for Eastern Europe*, CEPR, London, pp. 19-39.
- World Trade Organization. 1995. *Regionalism and the World Trading System*, WTO, Geneva.
- Yi, S. (1996), 'Endogenous formation of customs unions under imperfect competition: Open regionalism is good,' *Journal of International Economics*, 41, 151-175.

