

Nondiscrimination in GATT/WTO: was there anything to begin with and is there anything left?

T. N. SRINIVASAN*

Abstract: Nondiscriminatory treatment at its border of a like product coming from another WTO signatory (Article I on Most Favored Nation Treatment), and of domestic and foreign suppliers of like or similar products within its borders (Article III on National Treatment) are widely held to be the fundamental principles of GATT/WTO. Yet GATT included significant exceptions to nondiscriminatory treatment, for example, in its articles relating to customs unions and free trade areas, antidumping and safeguards. I argue that these exceptions have become dominant over time so that not much nondiscrimination remains in the global trading system. With the recent inclusion of services, intellectual property and trade-related investment measures, traditional GATT issues of tariff and non-tariff barriers at the border to market access have become less important compared to regulatory barriers inside the border. It is an open question whether nondiscrimination per se is a salient issue in thinking about multilateral disciplines in these new areas.

1. Introduction

Robert Hudec was an unusual scholar. Trained in law, he was an excellent economist among lawyers, an outstanding lawyer among economists, and equally at ease in engaging both groups in scholarly arguments. He has written extensively and incisively on international trade law and legal aspects of the (non) institution that laid out the rules of the international trade game, namely, the General Agreement on Tariffs and Trade (GATT) and its successor institution, the World Trade Organization (WTO).

The locus classicus for a critical assessment of the case against discrimination in the use of trade measures from economic, legal and political perspectives is Hudec's (1988) analysis. There is not much I can hope to add to this classic and will not attempt it in this paper. Many of his other writings on issues of discrimination, are reproduced in Hudec (1999a). His contributions to Bhagwati and

* Samuel C. Park, Jr Professor of Economics, Yale University, 27 Hillhouse Avenue, New Haven, Connecticut USA 06511. E-mail: t.srinivasan@yale.edu. I thank Kyle Bagwell, Wilfred Ethier, Bernard Hoekman, Philip Levy, and Robert Staiger for their helpful comments.

Hudec (1996) are also relevant. In contrast to a typical footnote-festooned article in law journals in which the ratio of footnotes (with a liberal sprinkling of Latin words!) to a turgid text often exceeds one by a large margin, Bob's articles are fun to read and the average footnote ratio in them is far less than one! One of my favorites is on GATT's legal system, originally published in 1970 and reprinted in Hudec (1999a). I would like to begin with it in my tribute to Hudec.

In this article, Hudec notes the contradiction in GATT between the complex and carefully drafted, 'with lawyerlike precision', articles on substantive obligations of the contracting parties, and enforcement procedures which 'present a front of ambiguity and uncertainty', and 'make no functional distinctions between breach of legal obligations and other grievances'. He viewed the contradiction as 'manifestation of a distinctive jurisprudence [that is] primarily the work of diplomats rather than their lawyers' (Hudec, 1999a: 17). A rationale for diplomats' rather than lawyers' jurisprudence in GATT was the perception that handling disputes over matters of international commerce was too important to be left to lawyers and judges. Hudec traces the skepticism about lawyers to a report in 1932 of the League of Nations' Economic Committee. After quoting the following from the report:

bodies composed of judges, who cannot be thoroughly well acquainted with all the details of economic life, and who are rather inclined to rely on criteria of pure law in judging cases in which situations of fact and technical considerations are of predominant importance, do not always appear to operate in a way satisfactory to the parties. Moreover, it appears... that the [International] Court itself is of the opinion that judicial settlement is not always the best way of settling disputes of an economic nature (League of Nations, 1932: 4)

he wryly remarks that: 'The talk about the crippled intellect of judges and lawyers makes interesting reading, but the rationale was not quite candid.' He then adds:

Most of the law's subject matter is economic, and a very large part of it is just as complicated as the subject matter of trade agreements. The real problem with lawyers and judges had been their failure to understand the need for compromise in these matters. The point (and the international lawyers of the day may well have missed it) was that these legal obligations were not meant to be enforced to the letter. (Hudec, 1999a: 21)

It is ironic that the diplomats' jurisprudence, to use Hudec's pithy description of the GATT legal system and particularly its 'political' dispute settlement mechanism (DSM), which, according to Hudec (1999b), functioned well, has been replaced in the WTO by an adversarial legal process.¹ Although the persons now

¹ Hudec (1999a: 17–41) discusses the negotiating history of the charter of the stillborn International Trade Organization (ITO) of which GATT was to have been the Commercial Policy Chapter. The first working draft of the ITO Charter submitted by the United States, 'outlined a three step procedure for dealing with legal questions. Complaints were to be investigated and ruled upon by the Executive Board, an eighteen member executive committee charged with conducting the day-to-day business of the

constituting the Appellate Body of the WTO's DSM cannot be characterized in the same way as the League of Nations report characterized the bodies composed of judges adjudicating economic disputes, the 'criteria of pure law' in judging economic disputes, which the report was critical of, seems to be ascendant in today's WTO. The deliberate and constructive ambiguity of diplomats' jurisprudence as incorporated in the GATT certainly opened up opportunities of legalistic battles of textual interpretation. Despite this, the political process of GATT, in which a defendant contracting party could veto the appointment of a panel for adjudicating the complaint against it, and both defendants and plaintiffs could veto the adoption of the findings of a panel if one were to be appointed, had strong incentives to nudge the parties to arrive at a compromise.² The WTO system, by making the appointment of a panel in effect automatic and requiring a consensus in the General Council for overturning the findings of the Appellate Body, has eliminated these incentives. Largely at the successful insistence of the United States (US), intellectual property protection was brought into the GATT, rather than having it remain with the World Intellectual Property Organization. US pressure was also behind 'legalizing' the DSM of the WTO. This has encouraged the demand for inclusion of labor standards and other issues into the WTO to take advantage of its perceived strong enforcement capabilities through its DSM.

Another consequence of the ascendancy of pure law jurisprudence is the possibility of judicial activism by the Panels and Appellate Body through interpretation of ambiguous GATT texts. There is some evidence that this is already happening in some of the decisions of the Appellate Body on some recent Article XX disputes,

ITO. Rulings of the Executive Board could be appealed to the Conference, the plenary body of all members. Rulings of the Conference in turn, could be appealed to the International Court of Justice. This final stage had one significant limitation. With only minor exceptions, appeals to the ICJ were to be allowed only "if the Conference consents" (US Department of State, 1948). In the debates over the draft, some objected to the 'consent' requirement on the grounds that the decisions of the Executive and the Conference might well be 'political' rather than objective. Others defended it essentially on the same grounds as the League of Nations Committee against adjudication of economic disputes by lawyers. Hudec describes the compromise that solved the 'consent' issue as follows: 'It was agreed that legal questions could be appealed to the ICJ as a matter of right. On the other hand, such appeals would have to take the form of a request by the ITO itself for an advisory opinion. The parties to the dispute would not appear as litigants before the court. In particular, the losing party would not be subject to a decree of the Court. Enforcement would be in the hands of the ITO, where, presumably, men of sound 'economic' judgment could take account of the 'economic facts' (Hudec, 1999b: 26). In the GATT, there was no provision for appeal to ICJ at all. Nor is there any in the WTO.

² On the other hand, the veto feature also encouraged disputants to ignore the dispute settlement process of GATT altogether and engage in threats and counter threats at an early stage, in effect drawing 'lines in the sand' and making a political compromise very difficult. In the WTO system, there is an incentive for a regime in power to impose a trade measure to influence domestic politics, fully aware that the measure would be found in violation of WTO rules by the DSM, but only after the desired outcome in domestic politics has been achieved. Removing the offending measure then would have no political cost to the imposing regime and inflict only a temporary loss to its trading partners. The recent imposition of tariff on steel imports prices to the mid-term elections to the US Congress and its later removal once the DSM ruled against it after the election is an example. I thank Wilfred Ethier for drawing my attention to these aspects of dispute settlement in GATT and the WTO.

for example, in the shrimp–turtles dispute. Judicial activism of domestic courts are kept in check by legislatures which can amend the constitution or repeal the laws interpreted too broadly by the judges. There is no analogue of the legislature in the WTO to act as a check on its Appellate Body.

In the rest of the paper, I describe (in Section 2) nondiscriminatory treatment as mandated in Articles I and III of GATT, followed (in Section 3) by exceptions to nondiscrimination in other articles and their continuing consequences, which together show that there is not much nondiscrimination left in the global trading system. Section 4 is devoted to a brief description of the role of MFN and reciprocity as means of self-enforcement of the GATT contract. In Section 5 I explore aspects of MFN and NT with a simple analytical model. Section 6 concludes by noting that, with tariff barriers on most goods becoming far less important in restricting market access, and the growing importance of services in international commerce, non-tariff barriers at the border and regulatory barriers inside the border have become far more pressing concerns than the discriminatory treatment that was the concern of Articles I and III.³

2. Nondiscrimination in GATT/WTO

Adverse discrimination

According to the Oxford Universal Dictionary of 1955, to discriminate is to ‘make or constitute a difference in or between; to differentiate’. Clearly, discrimination in this sense has no normative connotation. The dictionary goes further and defines, in a normative sense, ‘to discriminate against’ as ‘to make an *adverse* distinction with regard to’ (emphasis added) and interestingly, cites as an example, the discrimination (presumably by the United Kingdom ‘against certain imports from the United States’. Read with its meaning in the same dictionary, it follows that the use of the word ‘adverse’ in this context means that the discriminator is ‘acting against or actively hostile’ to the subject of discrimination, and the effect of discrimination is ‘unfavorable, injurious, calamitous’ to the subject.

Elimination of discriminatory treatment as a goal in GATT

The General Agreement on Tariffs and Trade (GATT) was a contract among its signatories, called contracting parties.⁴ Since tariff policy is a choice of sovereign

³ Although formally the General Agreement on Trade in Services (GATS) concluded at the Uruguay Round, includes the analogues of Articles I and III, namely, Articles II, XVI and XVII, the issues in services trade are different and more complex than those in goods trade. Whether discriminatory treatment is adequately addressed by these articles is an open question. See Mattoo (2000).

⁴ Not all signatories were nation states (e.g., Hong Kong). Their essential characteristic was that they constituted separate customs jurisdiction. In order not to confuse a signatory with a nation state in GATT documents, signatories are called ‘contracting parties’. The same documents, wherever something refers to all the contracting parties collectively, the phrase ‘Contracting Parties’ is written in capital letters instead of using ‘GATT’. The reason was to ensure that there was no hint of GATT being construed as an international organization (Hudec 1999a: 37–38). In this paper, by GATT I mean the version called

entities, a voluntary contract among them to restrain its exercise has to be mutually beneficial. Having signed (and ratified) the contract, the signatories commit themselves to abide by the terms of the agreement and not to engage in acts that either violate the terms or are not in conformity, even if they do not violate, with the expectation of mutual gains each party has about the behavior of others. In the context of GATT, an act by one contracting party has to be viewed as discriminatory if and only if the act is ‘unfavorable, injurious or calamitous’ in the sense of reducing the gains of one or more of the other contracting parties. Any such agreement, if it is to be meaningful, has to include a definition of an adverse discriminatory act and specify a remedial action. The determination of the adverseness or otherwise of an act from the perspective of the agreement will naturally depend on an identification of their interests that parties view as being promoted (and protected) by the agreement. The objectives of GATT, as explicitly stated by the original twenty-three contracting parties, are very broad: ‘raising standards of living, ensuring full employment and a large and steadily growing volume of real income and effective demand, developing the full use of the resources of the world and expanding the production and exchange of goods’ (GATT, 1994: 486). The Marrakesh Agreement establishing the World Trade Organization (WTO) reiterated these objectives and added the following: ‘expanding the production of and trade in goods and services, while allowing for the optimal use of the world’s resources in accordance with the objective of sustainable development, seeking both to protect and preserve the environment and to enhance the means for doing so in a manner consistent with their respective needs and concerns at different levels of economic development’ (ibid: 6). In agreeing to the GATT, the original contracting parties viewed themselves as ‘entering into reciprocal and mutually advantageous arrangements directed to the substantial reduction of tariffs and other barriers to trade and to the elimination of *discriminatory treatment* in international commerce’ (ibid: 486, emphasis added). Implicit in this is the presumption that reducing tariffs and other barriers to trade, as well as eliminating discriminatory treatment, were essential to the pursuit of the broad objectives of the contracting parties.

Principles of nondiscrimination in GATT architecture

The discriminatory treatment that the agreement seeks to eliminate or, more precisely, the nondiscriminatory treatment it mandates, is encapsulated in two fundamental principles. The first is the much celebrated ‘General Most Favored Nation’, or MFN, treatment for short, mandated by the agreement’s very first Article I. The second is the ‘National Treatment on Internal Taxation and Regulation’, or NT for short, which constitutes Article III. MFN treatment is about nondiscriminatory treatment by one party of all other contracting parties

GATT 1994 that was part of Annex 1A of the Uruguay Round Agreement. It is the original GATT of October 1947 with all its subsequent modifications.

with respect to customs duties and charges (i.e., ‘border’ measures) of any kind on products it trades with them. NT forbids discrimination by a party in favor of its domestic product with respect to domestic taxes and regulation (i.e., inside the ‘border’ measures). It is trivially obvious that to take cognizance of discrimination one has to limit the set of products such that any differentiation between products within the set with respect to border or inside the border measures would be deemed *prima facie* discriminatory. The text of GATT itself, reports of its working parties, panels on disputes, as well as the Appellate Body of WTO’s Dispute Settlement Mechanism, use several phrases such as ‘like products’, ‘similar products’, ‘directly competitive products’, and ‘competitive or substitutable products’ to mention a few, in an attempt to set such a limit. Depending on the meaning attached to each of the phrases, the associated sets could consist of a few or many products. To compound this problem, different articles of GATT use different phrases among the above list, thereby opening the possibility that such use was deliberate, that is the drafters of GATT in fact intended the article to apply to a few or many products depending on the phrase used.⁵

Several other articles of GATT refer to nondiscriminatory treatment. These include, in particular, Article II.1, VIII.1, IX.1, and XIII.1. Without in any way minimizing the significance of these articles in discussion of nondiscrimination in GATT/WTO, in the rest of the paper I will focus largely on aspects of MFN and NT covered by Articles I and III. GATT also included several exceptions to nondiscrimination, which taken together seriously weaken, if not altogether eliminate, the force of Articles I and III. I will return to these exceptions in Section 3 below.

Relevance of nondiscrimination in GATT in liberalizing market access

The signatories to GATT apparently did not view nondiscriminatory treatment as something to be of intrinsic value but only as an instrument for achieving the broad objectives they hoped to achieve with the agreement. In other words, they did not view nondiscrimination among contracting parties as epitomizing a principle of justice or fairness in some well-defined sense. This being the case, the argument sometimes advanced that such nondiscrimination in GATT is unfair because it implies ‘equal treatment of unequals’, whatever this means, is irrelevant. It is clear that the signatories viewed discriminatory treatment in

⁵ There is a large body of case law involving the interpretation of these phrases in GATT/WTO disputes (Bronckers and McNelius, 2000; Horn and Mavroidis, 2004; Hudec, 1999a: 365–367 and 2000; Neven, 2000). From my probably very biased perspective of an economist, much of the discussion in the panel reports and Appellate Body decisions appear to ignore the underlying economic issues in favor of textual analysis of GATT, very much in the tradition of the Brahmin ‘Bhashyakars’ (or interpreters) in Hinduism or of ancient rabbis of Talmud! Bob, who was clearly stung by the characterization of lawyers and judges as incapable of economic analysis in the League of Nations Report, surely would have chided me for my bias had he been with us!

international commerce, not as being unfair in some relevant sense (if any), but as subverting the broad objectives that signatories sought to promote through GATT.⁶ It raises the question of how precisely such subversion could arise.

A somewhat superficial answer to this question can be seen in the fact that the primary instrument of GATT for achieving the objectives is through liberalization of the access of each contracting party to the markets of other contracting parties. Any discrimination by one party in favor of a product originating in another party in its market for the product through border measures is seen as affecting the access of the parties not so favored to the same market.⁷ Article I prohibits such discrimination and uses the phrase ‘like products’ to limit the set of products within which prohibition applies. By the same token, an inside the border measure favoring a domestic product in its market by a contracting party is seen as affecting the access of all other contracting parties to the same market. Article III prohibits such discrimination among products within a set that is specific to each relevant paragraph of the article, if such discrimination is applied ‘so as to afford protection to the domestic product’ (GATT, 1994, Article III.1).

Analytical foundation for liberalization of market access

A deeper answer lies in the implicit analytical foundation for GATT’s emphasis on market access and its liberalization through multilateral tariff negotiations⁸ covered by Article XXVIII bis. This foundation is the proposition, often called the first theorem of neoclassical welfare economics, that a global competitive market equilibrium under free trade, loosely speaking, maximizes global welfare or, more precisely, is a Pareto optimum in the sense that any deviation from it, if it raises welfare of someone somewhere, it would do so only at the expense of someone else (Grandmont and McFadden, 1972). Thus, tariff or other barriers to trade that prevent global market integration, ipso facto, prevent the equilibrium in such segmented markets from being Pareto optimal. Given this proposition

⁶ Hudec (1999a: 227–250) provides very thoughtful analysis of the spuriousness and pernicious nature of fairness as a basis for trade laws and trade actions in the context of US trade policy. He rightly points out that ‘first, fairness is a matter that governments determine unilaterally: there are relatively few international agreements regulating the substance of such claims, and there is no recognized tribunal to adjudicate them in common law fashion. Fairness is largely what a government wants to call unfair (sic). Second, fairness claims call for unilateral concessions. No one pays to have unfair conduct corrected. To say that certain conduct is unfair is to say that the guilty party must correct it for that reason alone. In practice, this means that fairness claims can become a base for simply bullying others in search of greater commercial advantage’ (Hudec, 1999a: 227). His short paper on the political morality of trade negotiations (Hudec, 1999a: 215–225), though not about fairness, is also relevant since it raises the issue of whether pursuing legitimate objectives relating to health and environment would be undermined by the pursuit of the original objectives sought of GATT signatories through trade liberalization.

⁷ The relevant market from the perspective of Articles I and III, especially the latter, is not clearly delineated in GATT or in case law (Neven, 2000).

⁸ In fact, the eight successful rounds of multilateral negotiations under the auspices of GATT covered not only tariffs and other trade barriers but also rules governing trade.

as the foundation,⁹ the total elimination of barriers to free trade through successive reductions as *the* driving force of GATT/WTO makes eminent sense. However, the validity of this proposition depends on the satisfaction of some strong assumptions, including, in particular, that markets for goods, services and factors are perfectly competitive and there are no production or consumption externalities. Except for Article XVII on State Trading Enterprises, which recognizes the possibility of such enterprises not acting ‘solely according with commercial considerations’ (i.e., not behaving as competitive enterprises) and also their being monopolistic (such as import monopoly, marketing boards, etc.), GATT implicitly assumes that markets are by and large competitive. The ‘behind the border’ regulatory issues of competition policies, that are meant to ensure that markets remain competitive, and formulating multilateral disciplines for them, did not figure in the agreements concluding any of the eight rounds of GATT negotiations. The topic ‘Trade and Competition Policy’ was discussed in the first ministerial meeting of the WTO in 1996 in Singapore, and a working group was set up to study it. Subsequent attempts were made to include it in the negotiating agenda for the ongoing Doha round. These did not succeed. The General Council of the WTO decided on August 1, 2004 that no work towards negotiation on these issues will take place within the WTO during the Doha Round.

Externalities and GATT

Externalities in the precise sense of invalidating the Pareto optimality of a free trade competitive market equilibrium (FTCME) if unaddressed are also not explicit in GATT. However, the recognition in Article XX(b) that protecting human and plant life could justify departures from nondiscrimination should be viewed as, in fact, a recognition of failures either by producers or consumers of internalizing the external consequences to human or plant life of their actions in a FTCME. Another example is the agreement on Application of Sanitary and Phytosanitary Measures (SPM) concluded in the Uruguay Round. It elaborates rules for the application of the provisions of GATT, which relate to the use of sanitary or phytosanitary measures, in particular the provision of Article XX(b) and its chapeau.

Environmental externalities are well known. These are at the center of the debates on global warming, for example. The decision on trade and environment in the Uruguay Round to establish a committee on Trade and Environment open to all members of the WTO was a response to the claims that trade liberalization could conflict with environmental protection. It was also in part a recognition of

⁹ The analysis of Bagwell and Staiger (2002), which I discuss in Section 4, provides a related but distinct foundation, namely that tariff liberalization under reciprocity provides a means of escape to a mutually advantageous equilibrium from a non-cooperative one. This is consistent with the goals of the signatories of GATT, namely to raise standards of living through reciprocal and mutually advantageous tariff reduction.

the need for coordination of trade and environmental policies, without exceeding the competence of the multilateral trading system, which is limited to trade policies and those trade-related aspects of environmental policies which may result in significant trade effects for the members of WTO. The committee's terms of reference, *inter alia*, are to address 'the relationship between the provisions of the multilateral trading system and trade measures for environmental purposes, including those pursuant to multilateral environmental agreements; the relationship between environmental policies relevant to trade and environmental measures with significant trade effects and the provisions of the multilateral trading system' (GATT, 1994: 470–471). The committee has submitted several reports. At its last meeting, the General Council of the WTO in July 2004 simply took note of these reports without taking any further action.

Article XX(b), the agreement on SPM, and also Article III explicitly recognized the possibility of a protectionist misuse of the exceptions to nondiscriminatory treatment allowed under them and prohibited such misuse.¹⁰ Establishing such misuse of allowed nondiscriminatory treatment by a contracting party involves, in principle, showing that, first, there was discriminatory treatment; second, there was protectionist intent or *aim* behind it; and, third, that discriminatory treatment had the *effects* of hurting other contracting parties. The record of GATT/WTO jurisprudence on an 'aim and effects' test is mixed (Hudec, 1999a: 359–400).

For showing discriminatory treatment, it has to be shown, first, that the products being treated differently are among 'like' products (or any other phrase used in the relevant article for defining the set of products among which discrimination is disallowed by it). Second, in the GATT architecture (at least until the celebrated tuna/dolphin case) the only distinctions recognized were based on qualities of products themselves or other characteristics that indirectly affect product qualities, such as attributes of production processes or of the producer. This being the case, even if the discrimination is based legitimately on differences in products other than their qualities, it would violate Article III, for example. Showing protectionist intent, obviously involves an examination of policy makers' objectives, and, in particular, whether alternative less protectionist means were available for achieving their objectives. Showing effects of discriminatory treatment at the very least involves an analysis of the demand and supply behavior in the market for the products. Needless to add that establishing that nondiscriminatory treatment has, in fact, occurred is not straightforward.

10 For example, in Section 5, I consider a model in which a domestic consumption tax which discriminates in favor of a domestically produced import substitute, but which could be justified on the grounds of differential externality effects (for example, environmental pollution) between domestic and imported goods. The issue then becomes one of precluding the use of such a tax for purely protectionist (*i.e.*, to favor domestic substitute) purposes, while claiming it to be justified on environmental grounds.

3. Exceptions to nondiscrimination in GATT/WTO

Relevant GATT Articles

There are several articles of GATT that allow exceptions to nondiscrimination. As noted earlier, Article XX allows a contracting party to choose discriminatory measures, which are not, applied in a manner which would constitute ‘a means arbitrary and unjustifiable discrimination’ in order to achieve objectives listed in the article. Article III allows discriminatory use of internal taxes, laws and regulation as long as they are not ‘applied so to afford protection to domestic production’. Article VI, and the agreement in the Uruguay Round on its implementation deal with anti-dumping measures, which are by definition discriminatory, not only with respect to countries of origin of the product involved, but also, in principle, between exporters from the same country who are found to have dumped and those who are not. Article XIX on emergency action on imports of particular products, and the agreements on safeguards in the Uruguay Round relating to it, allow quantitative restrictions or quotas to be imposed.¹¹ Although Article XIII of GATT and Article 2, Paragraph 2, of the Safeguards Agreement explicitly prohibit the discriminatory application of safeguards measures by source, as Hudec (1988: 176–177) argued:

Where tariffs are concerned, there is a reasonably solid consensus about the core meaning of the most-favored nation (MFN) concept ... There is no such commonly accepted core meaning with regard to quantitative restrictions (QRs). There is agreement about one basic requirement: if a QR is imposed on imports of a given product, it must be imposed on imports of that product from all countries. Although this is a necessary condition, it is not a sufficient condition. Everyone agrees that the MFN principle requires something more in the case of QRs. It requires that the *manner* of applying the QR also be ‘nondiscriminatory’. Unfortunately, there has never been an agreed definition of exactly what this added requirement contains.¹²

Other articles permitting the imposition of QRs and hence, subject to the problem pointed out by Hudec, are the so-called balance of payments exceptions of Articles XII.1 and XVIII.9. Article XIV permits exceptions to nondiscriminatory treatment of QRs mandated in Article XIII in payments arrangements for current

¹¹ Administered protection measures such as voluntary export restraint measures and, of course, bilateral quotas such as those of the Multifibre Arrangements are, of course, discriminatory. However, they involve implicit compensation to the injured since they get to keep the quota rents. Because of this compensation feature, they make such measures less malign as compared to antidumping measures, which are discriminatory but do not involve compensation. See Ethier (2002). My reading of Ethier’s analysis is that some forms of discrimination combined with compensation for the discriminated against help in precluding the use of worse forms of discrimination. However, I do not see this as an argument for discrimination per se.

¹² Ethier, in a private communication, points out that to the extent MFN contributed to lowering of tariffs, it sowed seeds of its own destruction in the sense of making discriminatory non-tariff barriers more attractive, since tariff preferences are not attractive at low levels of MFN tariffs.

transactions in order to restrict imports. Since by definition Customs Unions and Free Trade Areas discriminate between their members and non-members with respect to tariffs, by not ruling them out, but explicitly allowing their formation subject to some admittedly stringent conditions laid out in Article XXIV, GATT included another permitted exception to nondiscrimination. Article XXI allows discriminatory treatment for national security reasons. Last, but not least, developing countries and least-developed countries have been exempted from almost all articles of GATT that mandate nondiscriminatory treatment. Although the provisions in GATT in favor of developing countries date back to 1955, Part IV on Trade and Development incorporated in 1964 and the so-called Enabling Clause of the Tokyo Round Agreement of 1979 contain the most important of the exemptions from GATT disciplines that developing countries have been allowed. These exemptions, under the rubric Special and Differential Treatment (SDT), have become, in effect, entitlements that developing countries wish to extend in the ongoing Doha Round.¹³ Thus, from its very inception, GATT's 'fundamental' principle of nondiscrimination was heavily diluted and compromised by exceptions.

*Have exceptions become the rule rather than proving it?*¹⁴

I will confine myself to three of the exceptions to nondiscrimination among the many discussed in Section A. The first relates to nondiscriminatory application of QRs. This has been explicitly violated during the GATT era, although the violations were not formally challenged. Agricultural trade has not been subject to disciplines of GATT regarding the use of QRs, export subsidies and non-discrimination almost since the signing of GATT in October 1947. Although the Uruguay Round Agreement on Agriculture began a process of reintegrating agricultural trade into GATT, it was fairly limited and far from complete.¹⁵ The 'July Package' adopted by the General Council includes an agreement 'to establish detailed modalities ensuring the parallel elimination of all forms of export subsidies and disciplines on all export measures with equivalent effect by a credible end date.' The end date is yet to be agreed. However, tariff-quota schemes will not be phased out. A complex tiered formula for overall reduction of aggregate measure of support has been agreed to, though total elimination of trade-distorting

¹³ The trade preferences and related discriminations span most of the developing countries in the WTO. The so-called G-90, consisting of 90 developing countries, is one such group.

¹⁴ Again, my *Oxford Universal Dictionary* defines: 'The exception proves the rule' as a legal maxim meaning 'Exception proves the rule in the cases not excepted.'

¹⁵ The attempt to convert various border measures into a single ad valorem tariff equivalent so that the resulting tariff could be bound and reduced led to a so-called 'dirty tariffication' process by which tariff equivalents were bound at absurdly high levels that were way above their applied levels. Reduction commitments from such high levels made a mockery of the commitments – even at the end of the period of implementation of these commitments, the bound levels would still be much higher than applied levels so that there was virtually no liberalization of agricultural trade in the Uruguay Round Agreement on Agriculture.

domestic support and of special and differential treatment are not contemplated. Thus, significant discrimination will remain in agricultural trade even after the conclusion of the Doha Round. Although QRs, such as 'voluntary export restraints' negotiated by the US and the European Union with Japan on its export of automobiles, were eliminated as part of the Uruguay Round Agreement, the possible use of QRs for addressing balance of payments (BOP) problems still exist, since provisions of Articles XII and XVIII permitting such use have not been repealed. In fact, India invoked Article XVIII when its QRs were challenged, even though it did not have any BOP problem. Only after the WTO's Appellate Body ruled against it did India eliminate QRs in 2001.

Trade in textiles and apparel has been exempted from GATT rules since 1961; the initial Short-Term Arrangement covering cotton textiles was quickly converted to a Long-Term Arrangement in 1962, and 12 years later this was expanded into the Multifibre Arrangement (MFA) in 1974, which covers trade in textiles made from almost all natural and man-made fibers! The MFA has been a particularly egregious exception to GATT rules: apart from being an outright violation of MFN, it also permits the use of bilaterally negotiated trade quotas (in violation of Article XI) on an item-by-item basis between each importer and exporter. One cannot imagine a worse way of segmenting and heavily distorting markets. Under the Agreement on Textiles and Clothing (ATC) in the Uruguay Round, MFA is scheduled to expire on 1 January 2005, after being in force for nearly three decades. However, the phasing-out of the quotas of MFA is back-loaded. Nearly half the quotas of MFA will remain in place until the moment before its expiry. Although the ten-year phase-out period was meant to enable the industry in developed countries which would be uncompetitive in a free trade regime to adjust, not much of an adjustment has taken place. The fear that a very competitive China may capture the market after the phase-out of MFA has already spawned attempts by a coalition of non-competitive producers in industrial and developing countries to seek a delay of the phase-out and also to use the special provision negotiated by members of the WTO in China's agreement of accession to the WTO to restrain its exports. Although Donald Evans, Commerce Secretary in the US government, has opposed delaying the phase-out, it is not entirely out of the realm of possibility that discriminatory treatment in some form will continue to exist even after the phase-out of MFA.¹⁶

The exceptions from MFN for customs unions (CUs) and free trade areas (FTAs) under Article XXIV of the GATT (and its updated version in the WTO), under Part IV of the GATT, relating to economic development adopted in 1965, and under the enabling clause of the Tokyo Round of MTN in 1979, 'Differential and More Favorable Treatment, Reciprocity and Fuller Participation

¹⁶ Ethier, in a private communication, points out that at least the injured DCs got compensated in part for quota rents they kept in the MFA. This being the case, replacement of MFA on its phase-out by other protective measures with no such compensatory features will hurt some DCs.

of Developing Countries', are serious, since any preferential and discriminatory trade agreements, were they to be found consistent with these clauses, could last indefinitely.

Any proposed CU or FT agreement was required to be promptly notified to the GATT for examination by a working party. In all, 98 agreements were notified under Article XXIV during the life of the GATT from 1947 to the end of 1994, including the most enduring of all, namely the European Economic Community (EEC) and the European Free Trade Area (EFTA). A further 11 agreements were notified by developing countries under the 1979 enabling clause. Working parties were established to examine virtually all agreements. While 15 working parties had not completed their examinations as of the end of 1994 and five did not report for various reasons, out of the 69 that had submitted their reports, only six explicitly acknowledged conformity with Article XXIV of the agreement, and this six does not include the EC or the EFTA, and only two of the six are still active! (WTO, 1995: 16). In fact, 'no agreement was reached on the compatibility of the Treaty of Rome [that established the EEC] with Article XXIV, and the contracting parties agreed that because 'there were a number of important matters on which there was not at this time sufficient information ... to complete the examination of the Rome Treaty ... this examination and the discussion of the legal questions involved in it could not be usefully pursued at the present time'. The examination of the EEC agreement was never taken up again' (WTO, 1995: 11).¹⁷

Whether or not a CU or FTA that is consistent with Article XXIV would have increased global welfare, it is abundantly clear that the procedures laid down for examining such consistency have not worked. The WTO report rightly concludes that:

[w]hile the purpose of the Understanding on Article XXIV is to clarify certain of the areas where the application of Article XXIV had given rise to controversy in the past, and particularly as regards the external policy of customs unions, it fell short of addressing most of the difficult issues of interpretation noted above. For example, no consensus emerged in the Uruguay Round Negotiating Group on GATT Articles concerning proposals made by several participants (notably Japan), to clarify the substantially-all-trade requirement. It is evident, therefore, that most of the problems that have plagued the working party process were not solved in the Uruguay Round. (WTO, 1995: 20)

Article V of the GATS, which corresponds in many ways to Article XXIV on goods trade, shares many of the unsolved problems of Article XXIV. In sum, the WTO articles and procedures as they are now are unlikely to succeed in the future in

¹⁷ Philip Levy, in a private communication, points out that the European Coal and Steel Community, the precursor of the EEC, was neither a CU nor an FTA. It clearly violated MFN. Its acceptance (or, more precisely, its non-rejection for its violation of MFN) by contracting parties of the GATT is the 'original sin' from which everything evolved subsequently!

resolving, any more than GATT articles and procedures did, the tension, if not outright contradiction, between discrimination which is an inherent feature of CUs and FTAs and the fundamental principle of nondiscrimination.

WTO (2004: 68–70) reports that the number of preferential regional trade agreements continue

to rise furthering the growth of preferential and discriminatory trade relations ... The impasse experienced at Cancún and the resulting uncertainty concerning the fate of the Doha Round has apparently precipitated the pursuit of RTA partners with the announcement of negotiations of several new RTAs and many more in the proposal stage ... in the Asian-Pacific region, where the expected finalization of several bilateral RTAs will soon grant Mongolia the status of the only WTO Member [among 148] not yet party to any RTA.

Developments in the global RTA landscape in 2003 show that all WTO Members are increasingly zealous in developing networks of preferential partners. While the promotion of free trade at a preferential level may exert leverage for openness and competitive liberalization in international trade relations thus benefiting the multilateral process, this strategy carries certain inherent risks. Such risks include the capacity of RTA partners to negotiate and administer multiple agreements with the attendant risk of diminished attention to the multilateral system; the creation of vested interests in FTA partners which will resist dilution of preferential margins at the multilateral level; and the impact of these agreements on third countries, for example through trade and investment diversion. The Committee on Regional Trade Agreements (CRTA), the body entrusted with verifying the compliance of RTAs with the relevant WTO provisions, continued its examination of RTAs in 2003. However, the CRTA made no further progress on its mandate of consistency assessment, due to long-standing institutional, political and legal difficulties. Since the establishment of the WTO, Members have been unable to reach consensus on the format, and the substance, of the reports on any of the examinations entrusted to the CRTA.

Of the original 23 contracting parties of GATT, nearly half were developing countries (DCs), and many of them still are. As of 1 October 2004, an overwhelming majority of the membership consists of DCs. However, DCs did not participate effectively in the GATT and the first six rounds of multilateral trade negotiations under its auspices. In the seventh Tokyo Round, which concluded in 1979, they participated in strength and with cohesion. Their experience in the GATT¹⁸ up to the conclusion of the Tokyo Round could be interpreted in two diametrically opposed ways. On the one hand, it could be said that from the Havana conference on, the developing countries have been repeatedly frustrated

¹⁸ See Hudec (1987) and Srinivasan (1998) for an extended discussion of the experience of developing countries in the GATT. On Special and Differential Treatment, see Hoekman (2004); Hoekman, Michalopoulos, and Winters (2004); and Keck and Low (2004).

in getting the GATT to reflect their concerns. Tariffs and other barriers in industrialized countries on their exports were reduced to a smaller extent than those on exports of developed countries in each round of the MTNs. Products in which they had a comparative advantage, such as textiles and apparel, were taken out of the GATT disciplines altogether. Agriculture, a sector of great interest to developing countries, largely remained outside the GATT framework. ‘Concessions’ granted to developing countries, such as the inclusion of Part IV on trade and development and the Tokyo Round enabling clause on special and differential treatment, were mostly rhetorical, and others, such as GSP, were always heavily qualified and quantitatively small. In sum, the GATT was unfriendly, if not actively hostile, to the interests of developing countries.

The other interpretation is that the developing countries, in their relentless but misguided pursuit of the import-substitution strategy of development, in effect opted out of the GATT until the Tokyo Round. The formal incorporation at the Tokyo Round of their demands for a differential and more favorable treatment, including not being required to reciprocate regarding any tariff concessions’ by the developed countries, triply hurt them: once through the direct costs of enabling them to continue their import-substitution strategies, a second time by allowing the developed countries to get away with their own GATT-inconsistent barriers (i.e., in textiles)¹⁹ against imports from developing countries, and a third time by allowing the industrialized countries to keep higher-than-average MFN tariffs on goods of export interest to developing countries. Instead of demanding and receiving crumbs from the rich man’s table, such as the Generalized System of Preferences (GSP) and a permanent status of inferiority under the ‘special and differential’ treatment clause, had they participated fully, vigorously, and on equal terms with the developed countries in the GATT and had they adopted an outward-oriented development strategy, they could have achieved far faster and better growth. The success of East Asia since the mid-sixties, and China and India since the eighties, suggests that the second interpretation is closer to the truth. Still, in the Doha Round DCs are insisting on the continuation and extension of SDT. It is likely that their demands will be conceded by the other members, thereby institutionalizing the inherently discriminatory SDT.

Several distinct issues get confounded in the arguments for and against SDT. First is the issue whether, within the same rule-based organization such as WTO, the rules should be contingent on a country’s status as developed, developing but not least developed, or least developed, given that the criteria for classification of countries has nothing to do directly with the rules? The answer is no as long as all members participate effectively in the making of rules.

¹⁹ Philip Levy correctly points out that these so-called ‘grey area’ measures, though they certainly violate the spirit of GATT, do not ‘formally’ violate GATT rules because the violation was not formally pronounced as such by a GATT body.

Undoubtedly, the impact of a tariff on an economy and its trading partners will, in general, depend both on the level of the tariff and also the market structure (i.e., whether it deviates from the norm of pure competition). It is often the case that the tariff structure (i.e., the vector of tariffs across commodities) in, say, developed countries, is such that the tariffs and other barriers on commodities that largely DCs export are higher (the problem of tariff *peaks*) than on other commodities, and they are also higher on processed goods than on raw or semi-processed material used in processing and which are again exported by DCs (the problem of tariff escalation). However, such tariffs are almost always on an MFN basis, so that their adverse impact on DCs is not from discriminatory treatment of DCs in comparison to other exporters, which would have been a violation of Article I. Certainly, there is a strong case for eliminating tariff peaks and tariff escalation, but it is not a case for SDT. I have argued that such peaks and escalation reflect, in large part, the fact that the DCs had ‘opted’ out of GATT for a long time. In the GATT negotiating process, reciprocity (i.e., exchange of tariff reductions by each part in return for similar reduction by others) is the norm. Having gotten an exemption from having to reciprocate, the DCs lost any bargaining leverage to get the other countries to lower peaks and avoid escalation and have to depend essentially on their good will. Asking for continuation of SDT is not of much help from this perspective.

In their very thoughtful analysis, Keck and Low (2004) distinguish the following five arguments that have been advanced in favor of SDT:

- Special and differential treatment is an acquired political right.
- Developing countries should enjoy privileged access to the markets of their trading partners, particularly the developed countries.
- Developing countries should have the right to restrict imports to a greater degree than developed countries.
- Developing countries should be allowed additional freedom to subsidize exports.
- Developing countries should be allowed flexibility in respect of the application of certain WTO rules, or to postpone the application of rules.

For arguments based on rights to be convincing, the rights have to emerge from a normative framework, be coherent in the sense of the existence of an agency whose duty it is to provide such rights, and have to have some universal values underpinning them. None of these hold for SDT as a right. The normative framework of GATT/WTO is the efficiency and Pareto optimality of *free trade* – SDT is, by definition, in conflict with it. It cannot be the duty of the WTO to provide such a right, since the WTO is simply an institutional framework for the conduct of trade relations among its members related to agreements they have concluded among themselves. Certainly, SDT has no intrinsic and universal value – whether it even has an instrumental value in promoting development is arguable. The second argument is, again, unconvincing. DCs should be interested in the removal of *all* barriers to market access and not just relative reduction of barriers to

their export compared to those of others. Since trade restriction is rarely the first best instrument for development, the third and fourth arguments are not valid.²⁰

The fifth argument is more convincing in that it asks for *flexibility* in the application of common rules rather than for different rules. It is reasonable to argue that implementation of rules, and regulations and reallocation of resources in response to trade liberalization do involve significant adjustment costs. Also, the administrative structure and institutions (including legal institutions) of DCs might not be adequate for implementing the rules and bringing about regulatory changes quickly. For these reasons, giving DCs a longer time to bring their trade policies and regulations in conformity with WTO rules and, above all, assisting them with resources to ease the burden of adjustment would be appropriate while formally repealing SDT.

It is evident that the exceptions to nondiscriminatory treatment that were built into GATT are no longer exceptions but in fact the dominant feature of the global trading system. It would not be surprising if a carefully estimated proportion of world trade in goods and services as of 2004 that is free of any discrimination turns out to be very small. Thus, discrimination is the rule and nondiscrimination the exception in the WTO.

4. MFN and reciprocity as self-enforcement devices

In Section 2, I suggested that the analytical foundation for the focus of GATT on progressive liberalization of market access with the end goal of barrier-free movement of goods and services is the first welfare theorem of neoclassical welfare economics, as extended to international trade, by Grandmont and McFadden (1972). This extension shows that under certain assumptions, a free trade competitive market equilibrium is Pareto optimal. An implication of Pareto optimality is that the global resource allocation and trade flows associated with it are efficient. Hudec (1988) lucidly discusses how discriminatory tariffs and quotas impose an efficiency loss. Both my argument and Hudec's in effect assume that the multilateral trade agreement GATT exists and focus on its rules and roles. The statement that avoiding the repetition of the inter-war experience with beggar-thy-neighbor trade policies, competitive devaluation, etc. was the prime motivation for the negotiations that led to GATT, though true, is not adequate. An analytical argument will not only explain why the agreement was in the mutual interests of the parties, but also why it has proved resistant to defections and remains stable over time, even though

²⁰ The Doha Round is often called the Development Round. This is unfortunate since it leads to the mistaken impression that once the Round is concluded to the satisfaction of all members of the WTO, the problem of development will be solved. Unfortunately, the problem is far more complex, and the barriers to trade and investment that developing countries face are not the dominant constraints on development. The relevant constraints are largely in domestic political economy.

it has no third party enforcement mechanism.²¹ In this section I will briefly note two contributions (Bagwell and Staiger, 2002 and Ethier, 2004), which provide such an argument and which also rationalize the pillars of MFN and reciprocity in the GATT.

Bagwell and Staiger distinguish three approaches to determining the problem that a trade agreement can be presumed to solve. In the first, which they call the traditional economic approach, the problem is a terms-of-trade or other externality from unilateral (i.e., non-cooperative) tariff setting by national welfare maximizing governments of countries which have market power. In the absence of a trade agreement, the externality would lead to a trading equilibrium, which is inferior, in the sense of being worse from every participant country's perspective, compared to an (cooperative) equilibrium associated with an appropriate trade agreement. In other words, a trade agreement is a means of escape from a Prisoner's Dilemma.²² In the second, called the political economy approach, the objective function of the government is generalized to include other considerations, particularly political considerations, besides national welfare. The generalization, while bringing in more realism, leads to the same conclusion as the traditional approach, namely that a trade agreement is a means of escape from a Prisoner's Dilemma. The third, the commitment approach, is different from the first two in that the inefficiency of the equilibrium without a trade agreement arises, not from an externality, but from the inability of governments to commit credibly to implementing contingent trade policies. Thus, a government that announces a policy of reforms including trade liberalization may be tempted to renege, if it is in its interests, once the private sector undertakes the actions consistent with reform. Anticipating such renegeing, the private sector will not undertake the actions, and the announced reforms will fail. However, if a trade agreement ties the government's hands, and this tying is credible because it is costly for the government to defect from the agreement, then credibility of tying, in effect, means credibility of the government's commitment to reform.

In the first two approaches, since the source of the inefficiency of a trade equilibrium in the absence of an agreement is the inefficiency from a terms-of-trade

21 Coneybeare (1987: 278–279) argues that multilateral agreements with MFN provisions are inherently unstable and would oscillate between multilateral arrangements and a reversion to bilateral and other such contracts. He finds confirmation for his argument in international tariff history. The reason for this oscillation is that 'a large-numbers MFN system will break down, in the absence of an enforcement mechanism against free-riders, but the ensuing contracting costs (and possibly predatory behavior) will create pressure for a collective return to MFN norms. What is needed is some efficient combination of multilateralism and bilateralism. Large numbers of bilateral negotiations may be time-consuming, but at least they do not create the same incentive for unconditional defection that is induced by a large-numbers MFN-public good game.' The instability problem is avoided by the reciprocity requirement in the analysis of Bagwell and Staiger.

22 Coneybeare (1987: part II) uses the title 'Prisoner's Dilemma Trade Wars' in describing various trade wars since the fourteenth century.

externality, the motivation for a trade agreement to avoid the inefficiency is clear. However, in the commitment approach, the source of inefficiency is basically domestic. As such, a trade agreement even if it contributes to its amelioration, unless it is the only means for amelioration, it is not necessarily the best means. The attempt of Bagwell and Staiger (2002: 4, footnote 4) to link the domestic credibility problem to a cross-country externality is not entirely convincing.

Turning to MFN, Hudec (1988: 177) points out that no part of the modern MFN concept, which in any case was developed solely with respect to tariff policy, has a very solid historical foundation. Besides the efficiency argument that an MFN tariff policy will ensure that imports are procured from the cheapest source, the other conventional argument in its favor is that, in its absence, any two countries would be reluctant to agree to an exchange of concessions in a trade agreement for fear that the benefits of concessions obtained for one party from another will be eroded if, in the future, the second party offers even greater concessions to a third party. This is the so-called ‘erosion of benefits or preferences’ argument. Horn and Mavroidis (2001) provide a comprehensive survey of the economic and legal aspects of MFN. Their description of the critical role played by MFN, combined with reciprocity in the Bagwell–Staiger analysis, is concise and illuminating.

Basically, in the GATT agreement, any renegotiation of tariffs under Article XXVIII and XXVIII bis is subject to the reciprocity requirement that renegotiations ‘maintain a general level of reciprocal and mutually advantageous concessions not less favorable to trade than that provided for in [the] agreement prior to such negotiations’. Bagwell and Staiger show that, while MFN insures against the conventional erosion of benefits of concessions in an agreement through opportunistic deviation of a subset of parties at the expense of others, it is not enough to eliminate fully such opportunism. However, MFN combined with reciprocity requirement will lead to a renegotiation proof agreement that is politically efficient, so that any deviation from it can raise one party’s utility – which allows for political considerations besides national welfare – only at the expense of another’s. It should be noted that the politically efficient tariffs that the renegotiation proof trade agreement sustains *are not zero*, but represent the minimum that can be supported by the agreement. Thus, the Bagwell–Staiger theory has nothing to say on whether the GATT style negotiation process conditioned by MFN and reciprocity is an efficient means for achieving a barriers-free global trading system.

In an extremely provocative and stimulating paper, Ethier (2004) argues that what he dubs as the ‘Received Theory’ of multilateral agreements, based solely on terms-of-trade externalities, is ‘irrelevant to actual multilateral trade agreements’ and ‘ACTUAL MULTILATERAL TRADE AGREEMENTS DO NOT PREVENT COUNTRIES FROM TRYING TO INFLUENCE THEIR TERMS OF TRADE’ (emphasis in the original). Moreover, in his view, more

sophisticated versions of Received Theory such as, for example, that of Bagwell and Staiger, have two implications:

- (i) Small countries will *never* sign on to trade agreements. If they did, they would surrender the use of trade policy for domestic objectives while receiving *absolutely nothing* in return.
- (ii) Large countries will negotiate *only* trade agreements that constrain terms-of-trade manipulation. Trade agreements that do not do this would, *for no reason*, surrender the use of trade policy for domestic objectives.²³

In reality, we observe exactly the opposite. (Ethier, 2004: 304)

Thus, according to Ethier, '*The Received Theory predicts what we do not see and cannot explain what we do not see and cannot explain what we do see*' (emphasis in the original). His own multicountry model deliberately does away with market power induced terms-of-trade externalities by assuming that no country has the ability to influence world prices. Governments have an objective function that reflects political externalities; these are basically two. First, 'political support' is 'more sensitive to the direct effects of government actions than to the indirect consequences [because] of bounded rationality [of the public]'; second, 'trade volumes influence political support independently, to some degree, from their implications for factor reward'. Negotiations take place in a succession of periods. Although he does not present a formal bargaining model – and the government objective function is a reduced form – still the model 'gives immediate and transparent explanations of gradual liberalization, reciprocity, MFN, and multilateralism'. He summarizes his argument as follows:

1. If governments, sensitive to political externalities, conduct trade policy as they claim to be conducting it, they will not liberalize unilaterally but will find reciprocal liberalization that does not go all the way to free trade desirable.
2. Reciprocal liberalization is valuable to a government because negotiated foreign liberalization allows that government credibly to claim credit, in an environment where many things change at once, for the good fortune experienced by exporters.
3. If reasonably close substitutes for each country's products exist elsewhere, reciprocal trade negotiations of any sort, that do not intend to go all the way to free trade, require that negotiating government's subsequent behavior toward third countries be constrained in some way.

²³ Prima facie it would seem that a small country would have an incentive to join a trade agreement, with MFN provisions, which includes large countries since it will enjoy the benefits of a reduction in its tariffs by a large country that raises the world price of exports of the small country. Ethier's argument does not appear to dispute this. It suggests that if *constraining terms-of-trade externalities* is the *sole* reason for a trade agreement, small countries, whose trade policies create no such externalities by definition, and who might wish to use such policies for domestic objectives, would have no reason to join such agreements.

4. MFN status is a useful way, and the simplest way, to constrain negotiating governments' subsequent behavior toward third countries. And, if used enough, it works, *not* because of the direct assurance it offers individual countries, but because it functions as an externality.
5. Multilateralism can internalize this crucial externality. Thus it is appropriate that Article I of the GATT establishes nondiscrimination as a *basic* principle, instead of defining it as a subject for negotiation.
6. If trade agreements do feature MFN, negotiations must, sooner or later, be multilateral in any event. (Ethier, 2004: 317–318)

Horn and Mavroidis (2001) are right in their assessment that:

The models of Bagwell and Staiger, and of Ethier, are rather different and emphasize rather different aspects of MFN. The former shows how the *sole* role of trade negotiations is to address international externalities that go through terms of trade, while Ethier only focuses on political externalities that do *not* go through terms of trade. Interestingly, both approaches suggest a positive role for MFN, in both cases when complemented with another salient feature of the trading system. In the case of Bagwell's and Staiger's analysis, MFN is complemented by reciprocity: MFN and reciprocity jointly imply that there need not be any negative externalities from bilateral renegotiations. Ethier, on the other hand, puts no restrictions on negotiations other than MFN, but instead argues that multilateralism is what prevents MFN from being eroded through concession diversion.

Both analyses, in essence, are about characterization of equilibrium: to the best of my reading, there are no real disequilibrium dynamics in either.²⁴ However, real negotiations in WTO have dynamics of their own, with suggestions of brinkmanship as a strategy.²⁵ The processes of first agreeing on a framework for negotiations, then on their modalities, and eventually to actual negotiations, are not easily modeled and, again to the best of my knowledge, have not been. It is not easy to discern any 'convergence' to an equilibrium property in these processes.²⁶ Although I have dismissed associating notions of fairness with MFN and reciprocity, perhaps it is the perception of fairness of both in some vague sense and the belief that insisting on both will ensure that whatever be the end outcome

24 Many of the game theoretic analyses of GATT, like most such analyses, implicitly or explicitly assume 'common knowledge', an assumption that is far from reality.

25 In the GATT ministerial meeting that launched the Uruguay Round at Punta del Este, Uruguay, at 6 p.m. the day before the meeting was scheduled to end, no agreement had been reached. The US delegation announced with great fanfare that they had their bags packed and would depart for the US the next morning whether or not there was an agreement. This threat seemed to have worked – agreement on all issues was reached by 4:30 a.m. the next morning (Srinivasan, 1998: 31).

26 Preferential agreements are said to be of help in persuading trade negotiators that their failure to conclude a multilateral agreement will unleash regional agreements. For example, it is claimed that the North American Free Trade Agreement helped in concluding the Uruguay Round Agreement and the plethora of agreements now in place would help an agreement in the Doha Round. Although politicians seem to hold this view, there is no convincing evidence for it..

of these unbelievably complex and inexplicable processes, it will turn out to be fair for all participants.

5. MFN and NT once again

The following simple model is meant to illustrate a fundamental problem with the exceptions to nondiscrimination in Articles III and XX(b). It goes beyond the relatively easier one to solve of defining more precisely the meaning of 'like' or 'similar' products. There are two goods, x and y . Good x has one domestic and two foreign, sources of supply. Good y has just two, one domestic and one foreign, sources of supply. Let the social welfare function of the economy be additively separable in utility of *private* consumers and *social* disutility, possibly from environment pollution it creates, from private consumption of good x . More specifically, social welfare W is given by:

$$W = U(X_D + X_{1F} + X_{2F}, Y) - V(X_D + \theta_1 X_{1F} + \theta_2 X_{2F}) \quad (1)$$

where X_D is consumption of domestically produced x and X_{1F} and X_{2F} are consumption from the first and second foreign sources of good x , respectively, and Y is the total consumption of good y .

It is to be noted that in private welfare *and* in social disutility the three sources of supply of good 1 are perfect substitutes, except that the marginal rate of substitution (MRS) in private utility between any two sources of supply is unity, while in social disutility, the MRS between domestic and the first (second) foreign sources of supply is $\theta_1(\theta_2)$.

The production side of the economy is described by the transformation function:

$$F(Q_x, Q_y) = 0 \quad (2)$$

where $Q_x(Q_y)$ is the domestic output of good $x(y)$. Assume the economy is a small open economy that faces an international relative price of unity for all goods in terms of the domestic good x and trade is balanced so that:

$$X_D + X_{1F} + X_{2F} + Y = Q_x + Q_y \quad (3)$$

By assumption consumers maximize private welfare U and ignore the social disutility $-V$. Then it is straightforward to note that a laissez faire free trade competitive equilibrium (LFTCE) will be characterized by:

$$\frac{U_1}{U_2} = \frac{F_1}{F_2} = 1 \quad (4)$$

where U_1 is the private marginal utility of total consumption of good x from all sources and U_2 is the private marginal utility of consumption of good 2. Equation (4) is the standard characterization of LFTCE by the equality of the

MRS in consumption (i.e., U_1/U_2) and the marginal rate of transformation (MRT) in production (i.e., F_1/F_2), and equality of their common value to the world relative price (i.e., 1). If good x is exported in equilibrium, clearly X_{1F} and X_{2F} would be zero, and, if it is imported, the consumer would be indifferent between dividing total imports in any fashion between the two sources of supply. Clearly, with free trade and no domestic taxes, both MFN and NT hold.

Consider now a social optimum. Now the social MRS between consumption of good x and good y will depend on what sources of supply are being used in positive amounts. In the optimum, production levels Q_x^o and Q_y^o will be the same as in the laissez-faire equilibrium. However, there are several possible consumption optima. Without loss of generality, assume $\theta_1 < \theta_2$.

Case I: In this optimum, good x is exported. At this optimum:

$$\frac{U_1(X_D^o, Y^o) - V_1(X_D^o)}{U_2(X_D^o, Y^o)} = 1 \tag{5}$$

where $-V_1$ is the social marginal disutility from consumption of good x .

$$X_D^o \leq Q_x^o \tag{6}$$

and

$$Y^o = Q_y^o + Q_x^o - X_D^o \tag{7}$$

For no imports of good x to take place, it is clear that θ_1 has to exceed 1.

Since consumers who ignore social disutility would be equating U_1/U_2 to the relative price of good x they face, implementation of this optimum as a competitive equilibrium will involve producers' facing the world relative price of unity for good x , and consumers facing $1 + (V_1/U_2)$. Thus, free trade plus a tax of V_1/U_2 on private consumption of good x will sustain this equilibrium. Once again, MFN holds and NT holds trivially, since only the domestically produced good x is consumed. This being the case, the consumption tax applies only on the domestic supply of good x .

Case II: In this optimum, part of the consumption of good x is from imports. It is straightforward to see that only the first foreign source of supply of imports will be used in this optimum, given $\theta_1 < \theta_2$. The relevant optimality conditions are:

$$\frac{U_1(Q_x^o + X_{1F}^o, Y^o) - V_1(Q_x^o + \theta_1 X_{1F}^o)\theta_1}{U_2(Q_x^o + X_{1F}^o, Y^o)} = 1 \tag{8}$$

$$X_{1F}^o + Y^o = Q_y^o \tag{9}$$

For sustaining this optimum as a competitive equilibrium the relative price in terms of good 2 of domestic good x facing producers has to be unity, while that facing

consumers has to be $1 + (\theta_1 V_1 / U_2)$, implying that consumption of good x has to be taxed to raise its consumer price by $\theta_1 V_1 / U_2$ over the producer price of unity. However, this is not enough to sustain the optimum, since at these prices there is no way to prevent consumers from choosing to split the imported component of their total consumption of good x in any way they like between the two foreign sources of supply. However, the social optimum requires them to use *only* the first source of supply. To ensure that this happens, the second foreign source of supply has to bear a marginally higher tax (or to receive a lower subsidy).²⁷ Such a tax or subsidy scheme will prima facie violate NT, although MFN treatment holds. For this violation to be within the permitted exceptions under Article III or Article XX(b), it has to be shown that the additional consumption tax on the supply of good x from the second foreign source, which is levied only to exclude it from the domestic market, is not being ‘applied so as to afford protection to domestic production’ of good x to a minimum consistent with maximization of social welfare. The effect of the discriminatory tax is certainly the exclusion of imports from the second source. Hence, without going into the ‘aim’ of the discrimination, there is no way to establish that discrimination is not protectionist.

The example was deliberately chosen for two reasons. First, it sidesteps the issue of ‘like’ product: from the consumer perspective, good x from any of the three sources is *identical*. From a social perspective, although they are not *identical*, they are nonetheless *perfect substitutes*. Second, the effect of a discriminatory consumption tax on one of the foreign sources of supply, by excluding it from the market, prevents it from competing with its domestic and the other foreign source of supply. The example clearly shows that the problem with the exception to nondiscrimination in Articles III and Article XX(b) is fundamental and not just lack of clarity about the content of phrases such as ‘like products’, ‘similar products’, and others. The fundamental problem is that there is no getting around an inherently difficult ‘aim’ test.

6. Conclusions

Several conclusions emerge from the discussion in the previous sections. First, although the principle of nondiscrimination as enunciated in Articles I and III of GATT is considered by most analysts as its very foundation, the exceptions to nondiscrimination in other articles of GATT considerably qualified its force ab initio.

²⁷ Note that an alternative scheme of setting an import quota of the amount X_{1F}^0 and auctioning it will run into the same problem. The quota holder has no incentive to import all of his quota from the first source. However, two quotas, an amount of X_{1F}^0 from the first source and nothing from the second will work. But this would be a violation of MFN in the application of quotas.

Second, the so-called ‘free trade’ areas and customs unions, for which Article XXIV allows (an exception to the MFN principle) under certain stringent conditions, and other preferential trade agreements (PTAs) have proliferated and continue to proliferate, so that *all* the 147 members of the WTO are either members also of a PTA or negotiating to become one. During the nearly six decades of the history of the GATT/WTO, only six such agreements were pronounced to be compatible with Article XXIV by a GATT working party, of which only two still exist. Interestingly, the European Union is yet to be pronounced to be compatible. Whether or not the PTAs are largely beneficial to their members, without inflicting serious damage to non-members, whether they promote ‘deeper integration’ of the economies, in some desirable sense, of members than non-discriminatory multilateral agreements, such as WTO, promote, or whether they are simply alternative means for powerful members of the WTO for extracting concessions from the less powerful on issues such as labor standards, investment, etc., on which there are no multilateral agreements, are open questions on which there is an enormous and growing literature. Regardless of the merits of PTAs, the fact that their proliferation has undermined, almost completely, the nondiscrimination principle of WTO is beyond doubt.

Third, the proliferation of PTAs, along with continuation of tariff quotas and Special and Differential Treatment of Developing Countries, and weaknesses regarding MFN and NT in GATS, have rendered nondiscrimination as almost obsolete in the WTO.

Fourth, there is no strong historical foundation for the modern concept of unconditional MFN. In game-theoretic approaches to trade agreements (Coneybeare, 1987; Caplin and Krishna, 1988), MFN norms could be counterproductive. As Caplin and Krishna (1988: 281–282) correctly point out:

If we view the bargaining process as yielding efficient outcomes, as for example with the Nash bargaining solution, then MFN simply limits the tools available to different countries, shifting in the utility possibility frontier. Hence the most positive aspects of MFN can only be illustrated when the bargaining process absent-MFN yields inefficient outcomes.

Problems of incentives created by MFN for free riding and delaying participation in negotiations by large traders and other negative aspects of MFN are extensively discussed in the literature, along with its positive efficiency promoting dimension. Although the work of Bagwell and Staiger (2002) and Ethier (2004) has restored an essential role (different in the two contributions) for MFN in achieving and sustaining a trade agreement, the fact that, as noted above, discrimination rather than MFN-based nondiscrimination is the reality in global trade suggests that the role that these authors attribute to MFN may no longer be that relevant.

Fifth, tariff and other barriers at the border have largely been reduced to insignificant levels thanks to the eight rounds of multilateral trade negotiation under

the auspices of GATT. Inside the border, regulatory measures have become far more salient with the conclusion of GATS and also with the increasingly successful attempts to introduce non-trade matters under the rubric of trade-related this or that. It is an open question whether nondiscrimination per se is an important issue in thinking about multilateral disciplines in these new areas.

I am sure, had Bob Hudec been with us, he would have authoritatively and illuminatingly analyzed the emerging issues. My reading of his work is that he was not wedded to any dogmatic adherence to any principle, be it of non-discrimination or of free trade and was skeptical of all forms of dogmatism. We all miss him, his intellect, and his charm.

References

- Bagwell, Kyle and Robert W. Staiger (2002), *The Economics of the World Trading System*, Cambridge, MA: MIT Press.
- Bhagwati, Jagdish and Robert Hudec (1996), *Fair Trade and Harmonization*, Vols I and II, Cambridge, MA: MIT Press.
- Bronckers, Marco and Natalie McNelius (2000), 'Rethinking the 'Like Product' Definition in GATT 1994: Anti-Dumping and Environmental Protection', in Cottier and Mavroidis 2000.
- Caplin, Andrew and Kala Krishna (1988), 'Tariffs and the Most-Favored-Nation Clause: A Game-Theoretic Approach', *Seoul Journal of Economics*, 1: 267–289.
- Coneybeare, John A. C. (1987), *Trade Wars: The Theory and Practice of International Commercial Rivalry*, New York: Columbia University Press.
- Cottier, Thomas and Petros C. Mavroidis (2000), *Regulatory Barriers and the Principle of Non-Discrimination in World Trade Law*, Ann Arbor: University of Michigan Press.
- Ethier, Wilfred J. (2002), 'Unilateralism in a Multilateral World', *The Economic Journal*, 112(179): 266–292.
- Ethier, Wilfred J. (2004), 'Political Externalities, Nondiscrimination, and a Multilateral World', *Review of International Economics*, 12(3): 303–320.
- GATT (1994), *The Results of the Uruguay Round of Multilateral Trade Negotiations: The Legal Texts*, Geneva: GATT Secretariat.
- Grandmont, Jean-Michel and Daniel McFadden (1972), 'A Technical Note on Classical Gains from Trade', *Journal of International Economics*, 2(2): 109–125.
- Hoekman, Bernard (2004), 'Overcoming Discrimination Against Developing Countries: Access, Rules and Differential Treatment', Paper presented at the conference 'Role of the WTO System in the World Economy', Cornell Law School and Cordell Hull Institute, 9–10 July 2004, Paris.
- Hoekman, Bernard, Constantine Michalopoulos, and L. Alan Winters (2004), 'Special and Differential Treatment of Developing Countries in the WTO: Moving Forward After Cancún', *World Economy*, 27(4): 481–506.
- Horn, Henrik and Petros Mavroidis (2001), 'Economic and Legal Aspects of the Most-Favored-Nation Clause', *European Journal of Political Economy*, 17: 233–279.
- Horn, Henrik and Petros Mavroidis (2004), 'Still Hazy After All These Years: The Interpretation of National Treatment in the GATT/WTO Case Law on Tax Discrimination', *European Journal of International Law*, 15(1): 39–69.
- Hudec, Robert (1988), 'Tiger, Tiger in the House: A Critical Appraisal of the Case Against Discriminatory Trade Measures', *The New GATT Round of Multilateral Trade Negotiations: Legal and Economic Problems*, Boston: Kluwer.
- Hudec, Robert (1999a), *Essays on the Nature of International Trade Law*, London: Cameron May.

- Hudec, Robert (1999b), 'The New WTO Dispute Settlement Procedure: An Overview of the First Three Years', *Minnesota Journal of Global Trade*, 8(1): 34–56.
- Hudec, Robert (2000), '“Like Product”: The Differences in Meaning in GATT Articles I and III', in Cottier and Mavroidis 2000.
- Keck, Alexander and Patrick Low (2004), 'Special and Differential Treatment in the WTO: Why, When and How?', Staff Working Paper ERSD-2004–03. World Trade Organization, Washington, DC.
- League of Nations (1932), 'Procedure for the Friendly Settlement of Economic Disputes between States' (Official No. C.57.M.32.1932.II.B), at 4.
- Mattoo, Aaditya (2000), 'MFN and the GATS', in Cottier and Mavroidis 2000.
- Neven, Damien J. (2000), 'The Economic Evaluation of Protection Under Art III: Some Principles and Suggestions for the Definition of a Trade Relevant Market', in Cottier and Mavroidis 2000.
- Srinivasan, T. N. (1998), *Developing Countries and the Multilateral Trading System: From the GATT to the Uruguay Round and the Future*, Boulder, CO: Westview.
- US Department of State (1948), 'Suggested Charter for an International Trade Organization of the United Nations', Article 76; reprinted in US Department of State Publication 2598, Commercial Policy Series 93, United States Department of State, Washington, DC.
- WTO (2004), 'Annual Report 2004', http://www.wto.org/english/res_e/reser_e/annual_report_e.htm, Accessed October 12, 2004.
- WTO (1995), *Regionalism and the World Trading System*, Geneva: World Trade Organization.