

United States – Anti-Dumping Act of 1916
(Original Complaint by the European
Communities) – Recourse to Arbitration by the
United States under 22.6 of the DSU, WT/DS136/ARB,
24 February 2004: A Legal and Economic Analysis

ROBERT HOWSE

University of Michigan Law School

ROBERT W. STAIGER*

University of Wisconsin

1 Introduction

This arbitration on the level of countermeasures, pursuant to the Dispute Settlement Understanding (DSU) 22.6, originates from findings against the United States with respect to a provision of the 1916 Anti-Dumping Act, which permits a private cause of action for treble damages against an alleged dumping importer or producer where the dumping meets certain criteria including, notably, a predatory intention. An unusual feature of this case is that, apparently, the provision in question has never been used, at least against the European Community (EC), the complainant. In any case, the provision in question was found to be in violation of the WTO Anti-Dumping Agreement, both because it would impose measures on dumping other than and in addition to anti-dumping duties, and because in so doing, the procedural requirements of the Anti-Dumping Agreement would not be followed.

* We thank participants at ALI meetings where we presented this paper for helpful comments, particularly Alan Sykes, Frieder Roessler, and Gary Horlick.

2 Legal issues

Article 22.4 of the WTO DSU provides that, where a Member suspends concessions in response to the failure of the losing party to implement a dispute settlement ruling where that Member was the complaining party, “The level of suspension of concessions or other obligations authorized by the DSB shall be equivalent to the level of the nullification or impairment.” The *United States — Anti-Dumping Act of 1916* Arbitration addresses the meaning of equivalence between the violation of international law and the countermeasures that respond to the violation. It has generally been assumed that nullification or impairment refers to the trade impact on the winning Member of the violation, not the competitive relationship as such; thus, countermeasures are limited to the value of the trade foregone as a consequence of the continuing violation. (This assumption will be reexamined from an economic perspective later in this chapter.) In the 1916 Arbitration, the EC proposed countermeasures in the form of a “mirror” suspension, in relation to the United States, of the very obligation with which the United States continued to be in non-compliance. Thus, the availability under the US statute of a private right of action for treble damages against a foreign firm engaged in certain kinds of dumping behavior had been held to be contrary to the Anti-Dumping Agreement in a number of respects; the EC proposed, as a countermeasure, to amend its own legislation to make available to European companies a similar right of action for treble damages against US firms.

The arbitrator was required to decide whether such legal equivalence meets the requirements of Art. 22.4, regardless of a possible actual asymmetrical *trade* or *economic* impact. The arbitrator held, based largely on previous arbitration rulings, that the expression “level” in Art. 22.4 implied a quantitative limit on the trade or economic effects of the suspension of concessions or obligations; while a Member may take countermeasures in a non-quantitative form, suspending an obligation rather than a concession such as a tariff binding, nevertheless the economic or trade effect of the original violation creates a fundamental limit on such action. Holding that a fundamental objective of countermeasures under DSU Art. 22 was to induce compliance, and that punitive purposes were excluded, the arbitrator found that it could not endorse the proposed countermeasures of the EC, unless it had evidence that the trade or economic effect on the United States would not exceed the trade or economic effect on the EC of the continuing original violation.

This ruling is the opposite extreme of the ruling in the Foreign Sales Corporation (FSC) Arbitration (albeit based on the special rule for countermeasures in the SCM Agreement); just as in the FSC Arbitration (see Howse and Neven, 2004)¹ where the arbitrator largely dispensed with proportionality as a limit on countermeasures, and viewed considerations of actual trade or economic effect as dispensable, in the 1916 arbitration, the arbitrator appears to have taken a view of equivalence that is narrower than that warranted by the text of DSU 22.4 and the general international law jurisprudence on countermeasures.

First of all, with respect to the text of 22.4, the arbitrator focused on the word “level” without also considering its immediate context – the notion of “equivalence.” The word equivalence implies proportionality between two things that are not entirely commensurate or reducible in value to a common metric. If 22.4 involved measuring and comparing the dollar value of the original violation in trade and economic terms with the dollar value of the proposed countermeasure in those terms, then the drafters would more plausibly have used the expression “equal” rather than “equivalent.” One does not normally ask whether two quantities are equivalent to one another, but whether they are equal, less, or greater than the one or the other. Determining whether two things are equivalent, by contrast, may involve asking whether they are similar in value, even if they cannot (without arbitrariness) be measured quantitatively in a common metric.²

Second, as Cannizzaro notes, “normative countermeasures,” where equivalence or proportionality is achieved by suspending a symmetrical obligation, have often been recognized as consistent with the principles of state responsibility. In the *Air Service Agreement* case, the United States engaged in a symmetrical breach of an aviation treaty in response to France’s initial breach of the same obligations towards the United States under the treaty: the United States claimed that, despite France’s argument that the losses to France entailed by this response exceeded

¹ Howse, Robert and Neven, Damien, 2004. “United States – Tax treatment for ‘Foreign Sales Corporations’ Recourse to Arbitration by the United States under Article 22.6 of the DSU and Article 4.11 of the SCM Agreement (WT/DS106/ARB): A comment,” in Horn and Mavroidis (eds.) *The WTO Case Law of 2002*, The American Law Institute Reporters’ Studies, Cambridge University Press, 2004.

² This is a complex issue conceptually: on the relationship between “egalite” and “equivalence” in economic justice, see A. Kojeve, *Outline of a Phenomenology of Right*, tr. Frost and Howse (2000).

those suffered by the United States from the initial breach, “the services were equivalent in law and hence proportional, . . .”³

In the *Air Service Agreement* case, the arbitral tribunal ultimately decided that normative equivalence could be based upon the importance in principle of the provision breached to both parties, i.e. its legal importance, but the tribunal also said that quantitative equivalence might be relevant as well depending on the circumstances. This is Cannizzaro’s position, too. From this perspective, the Panel in the *1916 Arbitration* was not wrong to insist that the quantitative economic effect of the breach in relation to the countermeasures had to be considered as relevant to equivalence; rather, the Panel erred in suggesting that the economic effect of the breach must always be regarded as a limit on countermeasures, *regardless* of normative considerations, including the gravity of the breach. Again, this is the mirror image of the mistake of the FSC arbitrator, which threw out the window any consideration of the proportionality of the countermeasures to the economic effect of the original breach on the defendant.

The WTO obligation breached in the *1916* case was a particularly good candidate for a focus on normative equivalence, since the actual trade or economic effects of the private action for treble damages were so speculative; a US court had never in the past actually made such an award against an EC company, and, on the other hand, relatively few anti-dumping actions in the EC had been brought against US imports, making any economic or trade effects of the countermeasures quite speculative.

A further error of the arbitral Panel in its ruling was to confuse the issue of the limits on countermeasures with the issue of whether the purpose of countermeasures was to reestablish the pre-breach equilibrium between the parties, to achieve compliance, or to punish. The Panel was correct to reject “punitive” countermeasures as consistent with the DSU, and was also correct to identify the achievement of compliance as a fundamental objective of countermeasures (see Howse and Neven, 2004: the ILC Articles clearly establish that the achievement of compliance is the sole legitimate purpose of countermeasures (Art. 49.1), and the language of the DSU supports this understanding of countermeasures and state responsibility). The arbitrator was *not* correct, however, in suggesting that the rejection of *punishment*

³ E. Cannizzaro, “The Role of Proportionality in the Law of International Countermeasures,” *EJIL* (2001) 889–916, pp. 902–903.

as a purpose of countermeasures implied that normative considerations could not, in appropriate cases, result in countermeasures based on a *normative* equivalence, which might exceed, to some extent, the economic costs of breach. Thus, although the ILC Articles recognize achieving compliance as the goal of countermeasures, the understanding of proportionality in the Articles refers to the injury suffered by the retaliating state, *taking account of the gravity of the wrong and the rights in question*. It is thus clear that there is no necessary tension between the purpose of countermeasures as compliance (in contradistinction to punishment or “corrective justice”) and the consideration of normative factors in determining the meaning of proportional or “equivalent” countermeasures; it is not *exclusively* economic harm that is to be taken into account in the evaluation of whether countermeasures are equivalent, even if it would be inappropriate to throw out the window considerations of the extent of economic harm in favor of normative considerations, as did the arbitral panel in *FSC*.

3 Economic issues

We identify and evaluate two key economic questions that arise in the Arbitrators’ Report on the US Anti-Dumping Act of 1916. First, when one government introduces a measure that nullifies or impairs benefits under the Agreement for another GATT/WTO member, how is the level of nullification or impairment to be defined? And second, once defined, how should the level of nullification or impairment be measured? We consider each question in turn.

3.1 *How is the level of nullification or impairment to be defined?*

A first fundamental issue confronted by the Arbitrators in *US Anti-Dumping Act of 1916* is how to define the level of nullification or impairment. Within the GATT/WTO, a market access commitment is understood to mean a commitment to maintain “conditions of competition” between domestic producers and foreign exporters. Nullification or impairment then occurs when a government introduces a measure that alters the conditions of competition between domestic producers and foreign exporters from those that could have been reasonably expected based on an earlier negotiation.⁴

⁴ See, for example, Analytical Index: Guide to GATT Law and Practice, Volume 2, WTO, 1995, pp. 657–658.

Based on the definition of nullification or impairment given above, a natural definition of the *level* of nullification or impairment might be the *magnitude* by which the new measure altered the conditions of competition from those that could have been reasonably expected at the time of the earlier negotiation. If the new measure is, for example, the introduction of a new tariff, then according to this definition, the level of nullification or impairment is simply the magnitude of the new tariff, expressed perhaps in specific form or perhaps in ad valorem form: the new tariff alters the conditions of competition between domestic producers and foreign exporters by raising the cost of foreign exporters in delivering their product to the domestic market. However, the Arbitrators clearly reject this view (p. 18, 5.30). Instead, they adopt the view that the level of the nullification or impairment is the “trade or economic effect” of the US Anti-Dumping Act of 1916 on the EU (p. 16, 5.23). This reduces to the following rule, quoted by the Arbitrators from the Arbitrators’ Report in EC-Hormones: “We must estimate trade foregone due to the . . .” measure in question (p. 16, 5.24).⁵

The “trade effects” approach adopted by the Arbitrators on this first fundamental issue raises a couple of central questions from an economic perspective. The first central question is, accepting for the moment that the trade effect is the appropriate focus, how is the “trade foregone” to be defined? Is this to be defined as the trade volume foregone at the *original exporter price*, or instead as the *equilibrium* trade volume foregone? These two magnitudes can be very different. For example, in response to a new tariff, exporters might reduce their prices to stabilize volumes, so that the equilibrium trade volume might change very little; but the trade volume at the original exporter price could still drop significantly. Figure 1 illustrates the point. With the foreign

⁵ The Arbitrators emphasize (pp. 16-17, 5.24–5.28) that their trade-effects approach follows in line with the general approach adopted in earlier arbitrations. However, given the nature and application of the US Anti-Dumping Act of 1916, any possible trade effects would have had to take the form of a “chilling” or deterrent effect, and the Arbitrators considered but ultimately rejected the quantification of such effects as too speculative. Consequently, most of the Arbitrators’ discussion with regard to the level of nullification or impairment caused by the 1916 Act focuses on direct monetary payments made by European companies to the United States as a result of final legal judgements or settlements under the 1916 Act. We focus our evaluation on the general economic rationale for the trade effects approach, and for simplicity we ignore in our discussion the additional means of nullification or impairment that arose more specifically in the context of the 1916 Act (e.g. final legal judgments against foreign firms).

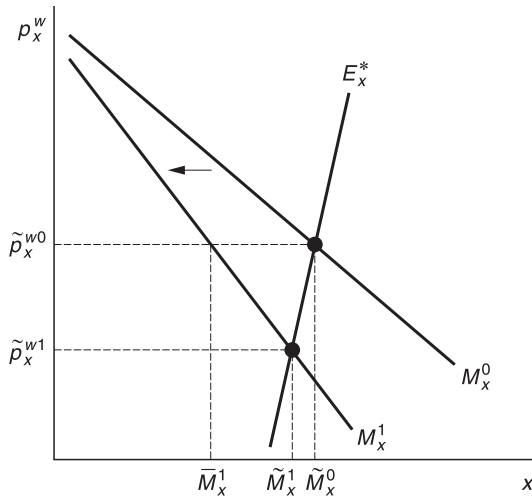


Figure 1

exporter (“world”) price of good x , p_x^w , measured on the vertical axis, and the quantity of good x measured on the horizontal axis, Figure 1 depicts an initial domestic import demand curve labeled M_x^0 and the foreign export supply curve labeled E_x^* . At the initial equilibrium (where the foreign export supply curve and the initial domestic import demand curve intersect), the (market clearing) exporter price is \tilde{p}_x^{w0} , and foreign exporters supply to the domestic country exactly the volume of good x that the domestic country wishes to import, \tilde{M}_x^0 . If the domestic country now imposes a new tariff on imports of good x , the domestic import demand curve shifts left to the curve labeled M_x^1 . At the original exporter price \tilde{p}_x^{w0} , the import demand of the domestic country is now the volume \tilde{M}_x^1 . Hence, the trade volume foregone at the original exporter price due to the new tariff is given by $[\tilde{M}_x^0 - \tilde{M}_x^1]$. On the other hand, according to the foreign export supply curve E_x^* , foreign exporters are willing to reduce their prices to stabilize volumes, and so the equilibrium (market clearing) import demand for the domestic country (determined in Figure 1 where the foreign export supply curve and the new domestic import demand curve intersect) is now the volume \tilde{M}_x^1 . Hence, the equilibrium trade volume foregone due to the new tariff is given by $[\tilde{M}_x^0 - \tilde{M}_x^1]$.

In the situation depicted in Figure 1, should it be concluded that very little nullification or impairment has occurred as a result of

the new domestic tariff, or rather that significant nullification or impairment has occurred? If the level of nullification or impairment is defined by the trade effect, the answer clearly would depend on which of these definitions of “trade effect,” (i.e. the equilibrium trade effect $[\tilde{M}_x^0 - \tilde{M}_x^1]$ depicted in Figure 1 or the trade effect at original exporter prices $[\tilde{M}_x^0 - \tilde{M}_x^1]$ depicted in Figure 1) is used.

The second central question raised by the trade-effects approach adopted by the Arbitrators is, why focus on the trade effect anyway? Once a departure is made from defining the level of nullification or impairment by the magnitude of changes in the “conditions of competition” (as in the magnitude of a new tariff), why not go all the way to the “welfare effect” on the complaining government? After all, the same trade effect could have vastly different welfare effects on a government, depending for example on the political pressure it faces or the adjustment costs shouldered by its firms and workers as a result of the changes in trade volume.

Hence, the trade-effects approach adopted by the Arbitrators raises two central questions: why focus on the trade effect?, and, how is the “trade foregone” to be defined? To begin to provide answers to these two questions, we first need to answer a prior question: What purpose is served in the WTO by the withdrawal of equivalent concessions? The Arbitrators’ Report considers this, and observes that the United States and the EU suggest that the authorized withdrawal of equivalent concessions can serve several useful purposes (p. 12, 5.3), including the restoration of the balance of benefits under the Agreement, and as a means to induce compliance. The Arbitrators observe that, whatever the complete list of purposes of withdrawals of equivalent concessions in the WTO may be, “. . . a fundamental objective of the suspension of obligations is to induce compliance.” (p. 13, 5.7). But the Arbitrators also suggest that the WTO rules dictate that the suspension of obligations cannot be punitive, and they indicate that the insurance that the suspension of obligations is not punitive is precisely that the withdrawal is “equivalent” (p. 13, 5.8).

According to the Arbitrators’ interpretation (p. 13, 5.8), then, the role served by the definition of “equivalent level of nullification or impairment” is to *ensure against punitive withdrawals*. Armed with the Arbitrators’ interpretation of the purpose served in the WTO by the withdrawal of equivalent concessions, we may now pose the question: why focus on the trade effect? The logic of the Arbitrators’ interpretation seems to suggest that one would need ultimately to quantify the *welfare*

impact of the change in conditions of competition on each government in assessing the level of nullification or impairment associated with the measure in question and determining whether a proposed withdrawal of concessions was or was not punitive. This logic suggests that a three-step analysis might therefore have been appropriate as a way to operationalize the Arbitrators' interpretation. To assess the level of nullification or impairment caused by the measure introduced by the defendant, one would focus on the trade impact of the defendant's measure and the welfare impact of this trade effect for the complainant. The three steps might run as follows:

Step 1: Measure the change in the conditions of competition brought about by the measure in question.

Step 2: Estimate the trade volume impact that is caused by the change in the conditions of competition. There is a question (as raised above) of whether this should be the equilibrium trade volume impact, or something else, but we will put this question aside for the moment.

Step 3: Factor in how the complaining government's welfare is affected by these trade volume changes. This will require knowledge of political pressures/dislocation costs suffered by the complaining government for any given change in trade flows.

In principle, this three-step analysis would provide an estimate of the welfare loss for the complaining government. And then, to ensure that the response is not punitive, the complaining government would be authorized to retaliate in such a way as to reduce the welfare of the defendant government by up to but not more than this amount. This would require a similar three-step analysis, but focusing instead on the trade impact of the complainant's proposed retaliatory measure and the welfare impact of this trade effect for the defendant government. Importantly, as Step 3 makes clear, by this logic it is not enough to quantify the trade effects alone — evidence on the relevant political pressures and dislocation costs must also be brought to bear in assessing the level of nullification or impairment.

But further reflection suggests that there are also alternative ways to operationalize the Arbitrators' interpretation, depending on what, exactly, is meant by "punitive." Does punitive retaliation mean imparting greater welfare loss on the defendant than was suffered by the complainant as a result of the measure in question? This first meaning of punitive is implicit in the three-step methodology described

just above, and it embodies a retaliation rule that resembles “. . . an eye for an eye, a tooth for a tooth.” Or does punitive retaliation instead mean imparting greater welfare damage on the defendant than the welfare increase that the defendant enjoyed from the introduction of the measure in question? Under this second meaning of punitive, the retaliation rule would be designed to return the defendant to its original welfare level that prevailed before the introduction of the measure in question: here, assessing the punitive nature of a retaliatory measure is all about quantifying the welfare impacts of various measures on the defendant, and there is no need to quantify the welfare impact of any measure on the complainant. One would still need to calculate trade effects associated with various measures at some level. And one would still need to know about the political/dislocation costs suffered by the defendant government from trade effects in different industries (i.e. the industry where the original measure was introduced, and the industry where the complainant’s retaliatory measures will be applied). But the calculations would be quite different under this second meaning of punitive than under the first.

The Arbitrators do not discuss the complications raised just above, which are associated with the ambiguity over the precise meaning of “punitive,” and simply assert that a retaliatory measure will not be punitive as long as it has no greater “trade or economic effect” on the defendant than the defendant’s original action had on the complainant (p. 16, 5.23). How should this assertion be interpreted?

We now suggest that, while the Arbitrators do not provide any further rationale for their assertion, it is possible to interpret the trade-effects approach as approximating a system of remedies that provides for “expectation damages” and facilitates a form of *efficient breach*. According to this interpretation, at a broad level, concessions negotiated within the GATT/WTO can ultimately be viewed as protected by an overarching “liability rule” — with the magnitude of the “liability” specified according to the trade-effects approach — under which a government may then escape from its commitments without the need to secure the permission of any adversely affected trading partner. In a wide variety of settings (e.g. Arts. XIX, XXIII, XXVIII), this in turn provides governments with a way to adjust the bargain when settlement negotiations fail. The broad interpretation we have just described has been developed in various forms by Sykes (2000), Schwartz and Sykes (2002), and Lawrence (2003) from a legal perspective, and Schwartz and Sykes also describe a possible political interpretation of

expectation damages.⁶ Our discussion below complements these earlier writings, by developing an economic interpretation of this point.

To motivate this economic interpretation, let us return to our earlier discussion of the definition of “punitive,” and consider its relation to compliance. As the Arbitrators record, the EU observes that “One could very well imagine that the best means to induce compliance would be to provide for punitive sanctions. However, the DSU does not allow for this possibility.” (p. 16, note 49). But why does the DSU disallow this possibility? A possible interpretation is that the DSU is designed to induce compliance, but not at “any cost,” i.e. not in all circumstances. Under what circumstances, then, is compliance to be induced? Perhaps only when it is “efficient” for the defendant country to comply. How do we know when this will be so? One possibility is to adopt the following interpretation of punitive: “A measure is *punitive* if it imposes a cost on the defendant that is so high that the defendant would comply even if it were not internationally efficient for the defendant to comply.” If such an interpretation could be made operational, then authorizing “non-punitive” retaliatory sanctions would ensure that compliance was induced only when it was efficient from an international perspective to do so. The key question is, How could such an interpretation be made operational?

To describe how, and to ultimately relate this interpretation back to the trade-effects approach to remedies embraced by the Arbitrators, we begin by observing that there is a third definition of “punitive” that we have not yet considered. According to this definition, punitive retaliation occurs when the claimant punishes the defendant so severely that the *claimant* ultimately enjoys higher welfare than if the defendant had not introduced the measure in question to begin with. Under this

⁶ See Sykes, Alan O., “The Remedy for Breach of Obligations Under the WTO Dispute Settlement Understanding: Damages or Specific Performance?” in Marco Bronckers and Reinhard Quick (eds.) *New Directions in International Economic Law*, Kluwer Law International (The Hague), 2000; Schwartz, Warren F. and Alan O. Sykes, “The Economic Structure of Renegotiation and Dispute Resolution in the World Trade Organization,” *Journal of Legal Studies* 31, 2002; and Lawrence, Robert Z., *Crimes & Punishments? – Retaliation Under the WTO*, Institute for International Economics (Washington, DC) 2003. Sykes (2000) and Schwartz and Sykes (2002) also offer a number of arguments for why the WTO might favor a liability-rule approach over the natural alternative of a “property-rule” approach, which would allow a government to escape from its commitments only after securing the permission of adversely affected trading partners. We refer the reader to those papers for further discussion of these and related issues.

third definition of punitive, the retaliation rule would be designed to return the claimant to its original welfare level that prevailed before the introduction by the defendant of the measure in question. At first glance, this third definition of punitive might sound strange, in that it builds from the view that the claimant actually enjoys the punishment. But of course, in the GATT/WTO, at some level this must be so: otherwise, since the GATT/WTO cannot coerce a claimant to punish (with a high tariff) if the claimant does not want to, the punishment would not be credible. So this third definition is not at all strange within the GATT/WTO context. And a little further reflection suggests that this third definition carries with it some appealing features from an economic perspective.

In particular, as long as the claimant remains below its best-response tariff, so that it would actually choose to raise its tariff to the full extent permitted by “non-punitive” retaliation under this third definition, the claimant is guaranteed to retain its original level of welfare (i.e. the claimant remains on its original “indifference curve”) as a consequence of (i) the introduction by the defendant of the measure in question, and (ii) the permitted retaliation by the claimant.

Suppose, then, that the domestic government understands the consequences of introducing a new measure that would nullify or impair the benefits that its foreign trading partner enjoys under the agreement (i.e. it understands the level of retaliation that this measure will induce). In choosing whether or not to introduce the measure in question — and if so, at what level — the domestic government will naturally make a choice that allows it to achieve its highest level of welfare. But as the permissible retaliation by the foreign government keeps the foreign government on its original indifference curve, the domestic government’s choice of whether to introduce the measure — and if so, at what level — amounts to making a choice that maximizes the domestic government’s welfare while preserving the level of the foreign government’s welfare. This, of course, describes a choice which is *internationally efficient*. Hence, under this third definition of “punitive,” we would have a system of remedies that provides for “expectation damages” and facilitates efficient breach.

Figure 2 illustrates the essential point in a hypothetical two-country two-good world. We denote by τ the level of import protection levied by the domestic government on imports of good x from the foreign country, and we plot τ on the vertical axis of Figure 2. Similarly, we denote by τ^* the level of import protection levied by the foreign

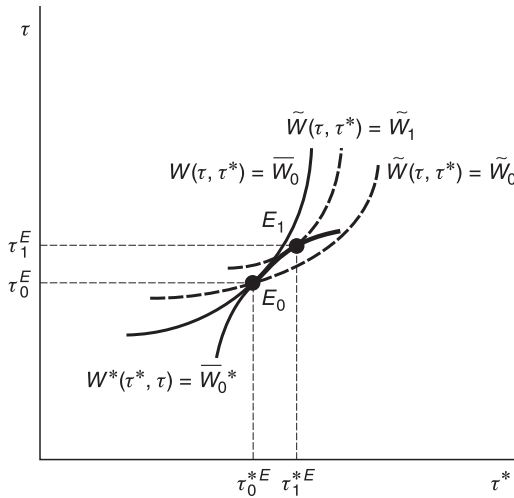


Figure 2

government on imports of good y from the domestic country, and we plot τ^* on the horizontal axis of Figure 2. We represent the welfare of the home government by the objective function $W(\tau, \tau^*)$ and the welfare of the foreign government by the objective function $W^*(\tau^*, \tau)$. For purposes of illustration, we suppose that the two governments have agreed to bind their tariffs at an internationally efficient pair of tariffs, denoted by (τ_0^E, τ_0^{*E}) and represented in Figure 2 by the point labeled E_0 .

Two features of the efficient tariff pair (τ_0^E, τ_0^{*E}) are reflected at point E_0 in Figure 2. First, the indifference curves of the two governments, labeled as $W(\tau, \tau^*) = \bar{W}_0$ for the domestic government and as $W^*(\tau^*, \tau) = \bar{W}_0^*$ for the foreign government, are tangent to each other at the efficient point E_0 . This is the defining feature of efficiency, because it indicates that there does not exist an alteration of the two tariffs which could make one government better off without hurting the other government. And second, the indifference curves are upward-sloping through the efficient point E_0 , reflecting the natural feature that, at the efficient tariff pair (τ_0^E, τ_0^{*E}) , each government would enjoy a welfare improvement if it were allowed to raise its own tariff unilaterally, while each government would suffer a welfare loss if the other government were allowed to raise its tariff unilaterally.

Now suppose that the domestic government experiences a “preference shock” – which might, for example, take the form of

an increased sensitivity to political pressures from its import-competing industry x — and that this shock leads the domestic government to place heightened value on the ability to raise its own tariff unilaterally. In Figure 2, this is represented by a shift in the domestic government's indifference curve through the point E_0 , to the *flatter* domestic indifference curve through E_0 represented by the dashed curve labeled $\tilde{W}(\tau, \tau^*) = \tilde{W}_0$. The indifference curve of the domestic government is now flatter through the point E_0 , because the domestic government would now accept a greater increase in the foreign tariff τ^* (a movement along the horizontal axis in Figure 2) in exchange for the ability to raise τ by a given amount (a movement along the vertical axis in Figure 2), owing to the heightened value it places on the ability to raise its own tariff.

A first point to notice in Figure 2 is that, as a consequence of the preference shock experienced by the domestic government, the point E_0 no longer represents an efficient pair of tariffs. This is reflected in the fact that the foreign government's indifference curve $W^*(\tau^*, \tau) = \bar{W}_0^*$ and the domestic government's indifference curve $\tilde{W}(\tau, \tau^*) = \tilde{W}_0$ are not tangent at E_0 .

The second point to notice in Figure 2 concerns the effect of a retaliatory rule, based on the third definition of punitive discussed above, on the domestic government's choice of whether, in light of its preference shock, to introduce a new measure in the form of a tariff which is higher than the bound rate τ_0 — and if so, at what level to raise its tariff. Recall from our earlier discussion of this retaliatory rule that, as long as the foreign government remains below its best-response tariff, so that it actually raises its tariff to the full extent permitted by “non-punitive” retaliation under this definition, the foreign government is guaranteed to retain its original level of welfare (i.e. the foreign government remains on its original indifference curve) as a result of (i) the introduction by the domestic government of the new measure, and (ii) the permitted retaliation by the foreign government.

Under this retaliatory rule, then, if the domestic government raises its tariff above the bound rate τ_0 , the permissible retaliation by the foreign government keeps the foreign government on its original indifference curve. In terms of Figure 2, this means that the foreign retaliatory response under this rule that is triggered by a tariff choice of the domestic government above the bound rate τ_0 is described by the bold portion of the foreign government's indifference curve $W^*(\tau^*, \tau) = \bar{W}_0^*$

above and to the right of the point E_0 . But understanding the retaliatory consequences of its tariff choice under this rule, the domestic government is then led to choose the tariff τ_1^E in Figure 2, thereby triggering the retaliatory response τ_1^{*E} by the foreign government permitted under this rule, and achieving its highest possible welfare level in light of this retaliatory rule, i.e. its highest possible welfare level along the foreign government indifference curve $W^*(\tau^*, \tau) = \bar{W}_0^*$. Under this retaliatory rule, then, the chosen act of “non-performance” (non-compliance) by the domestic government and authorized retaliatory response by the foreign government together result in the new efficient point E_1 , where a domestic government indifference curve labeled $\tilde{W}(\tau, \tau^*) = \tilde{W}_1$ is tangent to the foreign government indifference curve $W^*(\tau^*, \tau) = \bar{W}_0^*$. This is the essence of efficient breach.

By means of Figure 2, we have illustrated how this third definition of “punitive” could give rise to a system of remedies that provides for “expectation damages” and facilitates a form of efficient breach. Of course, this particular retaliatory rule would be virtually impossible to implement, because doing so would require knowing the indifference curve of the complaining government and setting up the retaliatory rule to correspond to this indifference curve. Knowing government preferences with such precision is of course impractical. But even if this information were somehow made available and the retaliatory rule described in Figure 2 could be implemented, we now suggest that its use would be problematic within the WTO contracting environment. As we next describe, the problem with this retaliatory rule is that the “efficient breach” that it would facilitate between any two contracting parties would have systemic *spillover* effects on other contracts between other contracting parties within the WTO system, leading indirectly to breach in these contracts as well. Hence, a retaliatory rule akin to that described in Figure 2 could introduce a kind of *multilateral instability* into the WTO contracting environment.

To understand the point, let us suppose that the situation depicted in Figure 2 occurs in a wider multilateral setting. In particular, let us suppose for purposes of exposition that the WTO membership consists of four countries, two domestic countries that export good y in exchange for imports of good x , and two foreign countries that export good x in exchange for imports of good y . Let us now refer to the domestic and foreign country depicted in Figure 2 as domestic country 1 and foreign country *1, and let us assume further that each is the

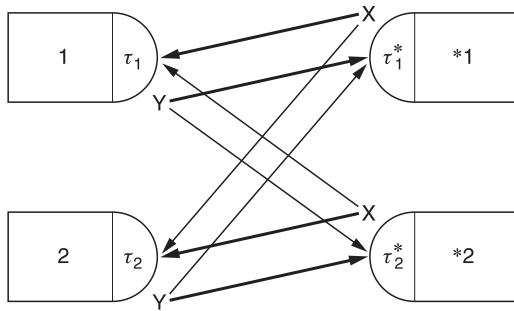


Figure 3

principal supplier of the other’s import good. We then refer to the remaining two countries as domestic country 2 and foreign country *2, and similarly assume that each of these countries is the principal supplier of the other’s import good. Finally, let us assume that each pair of principal supplying countries has negotiated a bilateral market access agreement that binds their respective MFN tariffs $((\tau_1, \tau_1^*), (\tau_2, \tau_2^*))$ at an internationally efficient level. Figure 3 illustrates the initial situation, with bold arrows indicating principal suppliers of a good and light arrows indicating lesser suppliers of a good.

Now suppose that, as in Figure 2, the domestic government 1 experiences a preference shock. And as in Figure 2, suppose that the system of remedies that provides “expectation damages” and allows for efficient breach is in place, and that under this system of remedies, the domestic government 1 breaches and the foreign government *1 retaliates so as to move from the point E_0 to the point E_1 in Figure 2. The key question is, Does the preference shock experienced by the domestic government 1 – and the breach and retaliation choices of domestic government 1 and foreign government *1 that this preference shock triggers – lead either the domestic government 2 or the foreign government *2 to choose to breach their contract also in light of the permissible retaliatory consequences of that breach? Generally, the answer to this question will be “Yes,” and in this way a retaliatory rule akin to that described in Figure 2 would introduce multilateral instability into the WTO contracting environment.

Intuitively, the international efficiency of the bilateral tariff bindings negotiated between principal suppliers depends in general on *all* the tariffs in the system, and so when the tariffs of domestic government

1 and foreign government *1, (τ_1, τ_1^*) , are allowed to change – as in the breach and retaliation contemplated in Figure 2 – the tariff bindings originally negotiated by domestic government 2 and foreign government *2, (τ_2, τ_2^*) , are generally no longer efficient, implying an opportunity for efficient breach of their bilateral contract as well. Hence, under this system of remedies, the implications of one breached WTO contract would in general be propagated throughout the WTO system and could in principle lead to a growing spiral of breaches.⁷

In light of the multilateral instability problem described just above, it might be tempting to conclude that the nature of WTO contracts makes provisions for breach unattractive, and that therefore remedies within the WTO should be designed to prevent breach (i.e. should be punitive). We next observe, however, that there is exactly *one* retaliation rule that could accommodate breach and yet in principle is not susceptible to the multilateral instability problem described above: this retaliation rule can be interpreted as akin to the trade-effects approach embraced by the Arbitrators.

Intuitively, the point is simple. Referring back to Figure 3, the spillover onto domestic country 2 and foreign country *2 that is generated by the breach and retaliation of domestic country 1 and foreign country *1 occurs as a result of the altered “third-country” trade flows between (i) domestic country 1 and foreign country *2 and between (ii) foreign country *1 and domestic country 2, and the implications of these altered trade flows for the market-clearing prices experienced by domestic country 2 and foreign country *2. Hence, if a retaliation rule could be designed in such a way as to prevent the

⁷ In principle, it might be possible that the resulting network of bilateral breaches could ultimately come to rest at a new internationally efficient point. But as a practical matter, the potential for instability seems real, and so we view these multilateral spillovers as a crucial fault associated with the remedies described in Figure 2. We also note that, by assuming that the tangency condition depicted in Figure 2 for the two-country case applies as well to an efficient point in our four-country example, we are implicitly assuming that the particular efficient MFN tariffs under consideration are “politically optimal” tariffs (see Bagwell, Kyle and Robert W. Staiger, “An Economic Theory of GATT,” *American Economic Review*, March 1999; and Bagwell, Kyle and Robert W. Staiger, *The Economics of The World Trading System*, MIT Press, 2002). More generally, as shown in Bagwell, Kyle and Robert W. Staiger, “Multilateral Trade Negotiations, Bilateral Opportunism and the Rules of GATT/WTO,” *Journal of International Economics*, December 2005, efficient MFN tariffs in a multi-country setting do not imply a tangency such as that depicted in Figure 2, but these (non-politically-optimal) efficient points suffer from a related kind of instability even in the absence of preference shocks.

alteration of third-country trade flows when one country breaches and a second retaliates in accordance with the rule, then breach under such a retaliation rule would not induce spillovers to third countries, and so would not create reasons for further breach of other WTO contracts. But we now observe that if retaliation is calibrated so that the breaching country's reduction in imports — when measured at original exporter prices — is *matched in volume* by the reduction in exports it experiences as a result of retaliation by its principal supplying country — also measured at original exporter prices — then there is no reason for trade flows with third countries to change, and hence no reason for the market-clearing prices experienced by third countries to change.⁸ In this general way, it may be seen that the trade-effects approach embraced by the Arbitrators has the attractive feature that it in principle *contains* breaches to the bilateral contracts directly involved (provided that foregone trade is measured at original exporter prices).

We have thus argued that an advantage of the trade-effects approach over *all* other approaches to the design of remedies that would permit breach is that the trade-effects approach provides a multilaterally stable remedy. A final question, though, is whether there are circumstances in which permitting breach under such a rule would be better than disallowing breach completely. After all, under the trade-effects approach, Figure 2 no longer applies, because Figure 2 illustrated the benefits of a different breach rule (and one that we subsequently argued would be multilaterally unstable). We now suggest that the trade-effects

⁸ This point is related to the seminal observations of Kemp, Murray and Henry Wan Jr., "An Elementary Proposition Concerning the Formation of Customs Unions," *Journal of International Economics* 6, 1976. The essential observation underlying the present discussion, that reciprocal tariff changes in the presence of MFN leave third countries unaffected, is shown in Bagwell and Staiger (2005). We also note that we have couched our arguments in broad terms, so that we may interpret essential features of the trade-effects approach as that approach is applied in a wide variety of settings (e.g. Arts. XIX, XXIII, and XXVIII). However, our multilateral stability arguments are made under the assumption that both the initial breach and the retaliation are non-discriminatory, and hence conform to MFN. This assumption fits well in the context of Arts. XIX and XXVIII. But discriminatory measures are often at the heart of Art. XXIII disputes, and Art. XXIII provides explicitly for discriminatory retaliation, so this assumption will often be violated in the context of Art. XXIII, weakening our arguments for the multilateral stability of the trade-effects remedy in this context. While a deeper exploration of the stability properties of the trade effects approach in a discriminatory setting would be interesting and worthwhile, we simply observe here that the economic appeal of the trade-effects approach will be strongest in settings where deviations from MFN are relatively minor.

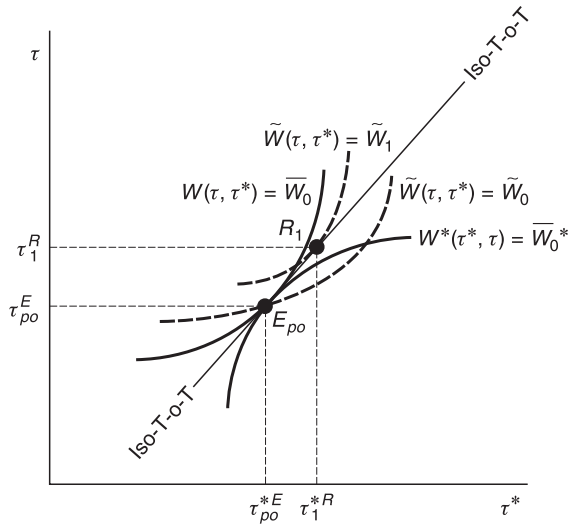


Figure 4

approach, in addition to generating a multilaterally stable remedy, provides a *local approximation* to the retaliatory rule illustrated in Figure 2 as long as governments begin from the efficient *politically optimal* tariffs.⁹ Hence, if initial negotiations have carried governments sufficiently close to the efficient politically optimal tariffs and provided that shocks to the status quo are sufficiently small, permitting breach in the presence of remedies designed along the lines of the trade-effects approach must be preferred to disallowing breach completely (say, with strongly punitive remedies).

Figure 4 illustrates the main points, returning again for simplicity to a two-country world. Again we depict the domestic tariff τ on the vertical axis and the foreign tariff τ^* on the horizontal axis. But now the efficient point from which the two governments are assumed to begin is the politically optimal point, labeled $E_{\rho o}$. As at the point E_0 in Figure 2 at the efficient point $E_{\rho o}$ in Figure 4, the indifference curves of the two governments are tangent. But at the politically optimal efficient point $E_{\rho o}$, the government indifference curves are also tangent to the locus of domestic and foreign tariff combinations that preserve the terms of trade between the two governments (see Bagwell and Staiger, 1999).

⁹ See Bagwell and Staiger (1999, 2002), for a development of the notion of politically optimal tariffs.

In Figure 4, this locus is labeled *iso-T-o-T*, and, as depicted, it must share the slope of the (domestic and foreign) government indifference curves at the efficient politically optimal point E_{po} .

The essential point from Figure 4 is that the *iso-T-o-T* locus, then, provides a local approximation to the indifference curves of *each* government around the politically optimal efficient point. So a retaliatory rule that traced out the *iso-T-o-T* locus would provide a local approximation to the retaliatory rule illustrated in Figure 2. This is illustrated in Figure 4, which depicts the point R_1 to which the two governments would move after a domestic preference shock as described above, if the domestic government faced the retaliatory rule described by the *iso-T-o-T* locus. As can be seen from Figure 4, for a sufficiently small preference shock, point R_1 can be brought arbitrarily close to an efficient point. Our final observation is then that the retaliation rule embraced by the Arbitrators *describes* an *iso-T-o-T* locus (provided that the trade effects are calculated at the original exporter prices). That is, the retaliation rule embraced by the Arbitrators is a *reciprocity* rule, under which domestic and foreign tariffs are altered in such a way as to bring about the same trade volume changes in each (domestic and foreign) market. But as shown by Bagwell and Staiger (1999), such reciprocal changes in tariffs *preserve the terms of trade*.

As a summary, we may return now to the two central questions raised by the trade-effects approach adopted by the Arbitrators: why focus on the trade effect? and, how is the “trade foregone” to be defined? We have provided a possible answer to the first question which suggests that the Arbitrators’ focus on trade effects has merit from an economic perspective: the trade-effects approach to remedies can be interpreted as approximating a multilaterally stable system of remedies that provides for “expectation damages” and facilitates a form of efficient breach. And within the context of our answer to the first question, the answer to the second question is now clear: to gauge the reciprocal retaliatory responses that would preserve the terms of trade between the defendant and the claimant, the trade effects should be defined and compared at the original (i.e. pre-tariff-change) exporter prices.¹⁰ That is, what should be quantified is the magnitude of the reduction in export volume that would be caused by the new measure at the original exporter prices.

¹⁰ See Winters, L. Alan, “Regionalism and the Rest of the World: The Irrelevance of the Kemp–Wan Theorem,” *Oxford Economic Papers*, 1997, for a discussion of the distinction between fixed equilibrium trade volumes and a fixed terms of trade.

In standard economic terms, this, in turn, is equivalent to quantifying the inward shift of the import demand curve at the original exporter price. From the perspective of the WTO, this can be interpreted as a market-access-based definition, because it quantifies the extent to which the alteration of the conditions of competition would have led to diminished export volume at the original exporter prices.¹¹

3.2 *How should the level of nullification or impairment be measured?*

Having developed an economic interpretation for the Arbitrators' approach to defining the level of nullification or impairment, we now turn to our second question, and consider how the level of nullification or impairment should be measured. Under the interpretation developed in the previous section, the level of nullification or impairment should be measured by foregone trade flows (and any direct economic impacts not captured by the foregone trade flows, e.g. legal judgments paid by exporting firms to the importing government).

As for measuring foregone trade flows, according to our interpretation in the previous section, the appropriate measure is a "market access" measure, i.e. a measure of the extent to which the alteration of the conditions of competition would have led to diminished export volume at the original exporter prices. This, in turn, is equivalent to measuring the inward-shift of the import demand curve at the original exporter price.

There is also a question of whether to include any "chilling effect" in measuring the foregone trade flows (p. 24, 5.64). The Arbitrators conclude that attempts to quantify such effects would be too speculative (p. 25, 5.69), at least for the present case. Although, in general, such effects may well be important, and indeed the existence of effects analogous to these have been empirically documented in the context of US anti-dumping law (see, e.g. Staiger and Wolak, 1994 and Krupp and Pollard, 1996), the Arbitrators' position on this nevertheless seems reasonable, for at least two reasons.¹² First, the 1916 Act applies

¹¹ For further discussion of the links between standard economic arguments and the GATT/WTO concept of market access, see Bagwell and Staiger (2002).

¹² See Staiger, Robert W. and Frank A. Wolak, "Measuring Industry-Specific Protection: Antidumping in the United States," *Brookings Papers on Economic Activity: Microeconomics*, 1994, pp. 51–118; and Krupp, Corinne M. and Patricia S. Pollard, "Market Responses to Antidumping Laws: Some Evidence from the US Chemical Industry," *Canadian Journal of Economics* 29(1), February 1996, pp. 199–227.

only where predatory intent can be established, and available evidence suggests that only in rare instances of dumping could a claim of predatory intent have been credibly made.¹³ Hence, given the rare circumstances in which the 1916 Act could be applied, it is unlikely that the mere existence of the Act could have a significant chilling effect on trade flows. And second, given the inherent uncertainties associated with any efforts to quantify chilling effects (as opposed to the effects of applied measures), it seems prudent to exclude consideration of chilling effects from efforts to quantify the *level* of nullification or impairment, though perhaps consideration of chilling effects could be the basis for a finding of the *existence* of nullification or impairment.

4 Conclusion: “Efficient Breach” and the law of the WTO

Our economic analysis shows how one can make sense of the arbitral Panel’s decision to limit countermeasures to an amount thought to represent the economic cost of the continuing breach to the plaintiff WTO Member, and the Panel’s corresponding rejection of the possibility of normative equivalence (albeit, in the Panel’s case, based on a confused notion of punitiveness). Further, there may be good policy reasons to adopt an approach to compliance with WTO obligations based on the notion of “efficient breach.” On this approach, it makes no sense to consider compliance as a superior outcome, if the breaching country can continue to breach while paying the full economic cost of its behavior.

The first set of legal issues, however, concerns whether in fact this approach is consistent with the law of the DSU as it is written. John Jackson has made a powerful case to the contrary.¹⁴ Jackson points to a number of provisions in the DSU, including 3.7, which states, “The first objective of the dispute settlement mechanism is usually to secure the withdrawal of the measures concerned . . .,” provisions that provide for the withdrawal of the measure as the normal recommendation of the Dispute Settlement Body (DSB) and that provide that

¹³ See, for example, Shin, Hyun Ja, “Possible Instances of Predatory Pricing in Recent Antidumping Cases,” in Robert Z. Lawrence (ed.), *Brookings Trade Forum*, 1998, pp. 81–98.

¹⁴ J. H. Jackson, “The WTO Dispute Settlement Understanding—Misunderstanding on the Nature of Legal Obligation,” 91 *American Journal of International Law* 60 (1997).

compensation and suspension of concessions are not preferred “to full implementations of a recommendation [to withdraw the offending measure].” (DSU 22.1) Moreover, Jackson argues that both the continuing surveillance of implementation that is required by DSU 22.8, and the statement of “no obligation to withdraw the measure” (DSU 26:1(b)) in non-violation (nullification or impairment) findings, are further evidence of the continuing nature of the legal obligation under the DSU to withdraw the measure regardless of the existence of countermeasures for failure to do so.

Sykes counters Jackson as follows: “If the objective of the system is ‘usually’ to secure withdrawal of the offending measure (Art. 3.7), it is impliedly not always the objective. This passage seemingly leaves doors open to other options. The fact that the Panels and the AB must recommend withdrawal of the offending measure (Art. 19:1) says nothing about the existence of an obligation to follow a recommendation, in preference to paying damages as an alternative... And if compensation and retaliatory measures are not ‘preferred’ (Art. 22:1), neither does the text expressly say that they are legal or that a Member who elects compensation or retaliation is in violation of the rules.”¹⁵ Sykes goes on to note that the obligation of surveillance of implementation is consistent with a “damages” approach, since one would want to make sure that continuing damages and their level are consistent with the continuing non-implementation. As for the fact that there is no duty to withdraw the measure in the case of *non-violation* nullification and impairment, Sykes argues that this does not necessarily imply there *is* such a duty in the case of violation; it could simply be an instruction to the dispute settlement organs that they can dispense with a *recommendation* of withdrawal of the measure.

Sykes’ analysis does show that the text of the DSU is ambiguous on the nature of legal obligation. However, the text of the DSU should not be read in isolation. It must be read in light of any other rules of international law applicable between the parties, and these include the customary international law of state responsibility. There are

¹⁵ A. O. Sykes, “The Remedy for Breach of Obligations under the WTO Dispute Settlement Understanding: Damages or Specific Performance?” in M. Bronckers and R. Quick, eds., *New Directions in International Economic Law: Essays in Honour of John H. Jackson* (The Hague: Kluwer, 2000), pp. 347–357.

unquestionable aspects of the WTO dispute settlement system that reflect a “contracting out” of the law of state responsibility, now codified in the ILC articles — for example, the unavailability of reparations in dispute settlement to compensate the plaintiff for retrospective harm. But given there is ambiguity in the DSU on the nature of legal obligation, it would be appropriate to resort to the interpretation that is consistent with the customary international law of state responsibility, and that is clearly Jackson’s interpretation. *Pacta sunt servanda* is fundamental to that law: countermeasures (as opposed to retrospective compensatory damages) are aimed at securing compliance, and they are not a substitute for it. Contracting out of custom should be clear and unambiguous.

There is perhaps one avenue of response to this objection that is open to Sykes: he could argue that the rules of state responsibility are beside the point, since the nature of the *primary* obligations in the WTO treaties is that they are obligations to compensate those harmed when the state in question does not act in conformity with the provision in question. This view of the primary obligations in the WTO treaties, however, seems directly contradicted by the explicit general statement of the nature of WTO obligation, contained in XVI:1 of the *WTO Agreement*: “Each Member shall ensure the conformity of its laws, regulations and administrative procedures with its obligations as provided in the annexed Agreements.”

Moreover, when the drafters of the WTO treaties wished to express an obligation in terms of a duty to compensate for a given course of conduct, they knew how to do so. Likewise, if for example, one considers provisions on expropriation in investment provisions of trade agreements or in BITs, the drafters generally make clear where expropriation is per se illegal on the one hand, and where it is legal, conditional on compensation, on the other.

In sum, reading the provisions of the DSU in light of the customary international law of state responsibility, Jackson’s view of the nature of the compliance obligation in the WTO is the stronger one. Nor are primary obligations under the WTO treaties in almost all cases obligations to compensate for engaging in the course of conduct addressed by the obligations. Yet these observations about *lex lata* in no way close the issue of whether as a matter of policy Sykes is right about “efficient breach” being a preferable approach to compliance. Our own economic analysis in this paper is consistent with Sykes’ preferred policy direction. In addition, Esserman and Howse (2003) and

Howse and Nicolaidis (2003)¹⁶ have argued that there may be cases that try the political legitimacy of the dispute settlement system, because the existing norms are too general or too controversial, or simply not written well enough to enable a legitimate outcome, where the finding of violation is acceptable (if obviously undesirable) to the loser. In these situations it might be desirable to have a mechanism where in effect, in relation to the findings of the dispute settlement organs, a losing party's obligations are actually waived in return for the payment of appropriate compensation to the winner (and here damages might well be a better form of "payment" than countermeasures). As conceived by Esserman and Howse, however, such a solution should still require consent of the winning party, which ought not be forced to give up its right to demand implementation.

A wholesale move away from *pacta sunt servanda* to "efficient breach" as the default approach to compliance in the WTO system would need to be based on a consideration of important systemic factors to:¹⁷

- 1) There are many different kinds of WTO obligations. Are some of these of a "public goods" nature, where the cost of non-compliance cannot be fully internalized by "payments" of the individual party or parties who are continuing to violate? (MFN, arguably, as well as transparency obligations).
- 2) Apart from the FSC arbitration, where (under the specialized rules of the SCM Agreement) an arbitral panel suggested that a party could demand countermeasures that represented economic harm to *all* WTO Members from a continuing violation (not just the complainant), the general approach is that countermeasures are awarded corresponding to the prospective economic harm to the winning party. In calculating a quantum for efficient breach, should one take into account only the economic cost to the complainant countries and ignore those to other countries who are not parties to the proceedings? Under the current rules of state responsibility,

¹⁶ Susan Esserman and Robert Howse 2004, "The WTO on trial," *Foreign Affairs*, 82(1), 130–140. Robert Howse and Kalypso Nicolaidis, (2003), "Enhancing WTO legitimacy constitutionalization or global subsidiarity?" *Governance*, 16(1), 73–94.

¹⁷ We have benefited from the illuminating discussion of related and similar considerations in John H. Jackson, "Status of WTO Dispute Settlement Reports: Obligations to Comply or Option to 'Buy Out'?" *American Journal of International Law*, January 2004, 109–125.

while it is improper to assess countermeasures against the losing party based on the harm to all WTO Members, nevertheless injured non-parties can still free ride as it were on the winning party's compliance efforts, since under the current rules, there remains an underlying obligation to bring oneself into conformity. The systemic implications of a losing Member buying itself out of its duty to comply with a payment based on harm to the complaining state(s) alone need to be carefully considered.

- 3) Are there equity issues raised by the fact that different WTO Members will have vastly different capacities to buy themselves out of obligations? What if the value to the breaching state is high but nevertheless it cannot afford to pay the cost that it is incurring on the victim state?