The European Journal of International Law Vol. 30 no. 1

Last Mile for Tuna (to a Safe Harbour): What Is the TBT Agreement All About?

Petros C. Mavroidis*

Abstract

The WTO Agreement on Technical Barriers to Trade (TBT Agreement) aims to tame non-tariff barriers, the main instrument segmenting markets nowadays. Some of the terms used in the TBT Agreement to flesh out the commitments undertaken are borrowed from the General Agreement on Tariffs and Trade (GATT), and some originate in the modern regulatory reality as expressed through standard-development organizations. The TBT Agreement does not share a copycat function with the GATT though. Alas, the World Trade Organization's Appellate Body, by understanding words as 'invariances' – for example, interpreting them out of context (without asking what is the purpose for the TBT Agreement) – has not only exported its GATT case law but also misapplied it into the realm of the TBT Agreement, and ended up with significant errors. This article explains why the current approach is erroneous, and advances an alternative understanding, which could help implement the TBT Agreement in a manner faithful to its negotiating intent and objective function.

1 Tuna Have Been Swimming in the GATT/WTO Waters since 1989

The tuna saga, a series of disputes between Mexico and the USA regarding the commercialization of tuna in the US market, has been ongoing for almost 30 years now. It has earned its place among the legendary *US – Softwood Lumber*, *EC – Bananas*,

* Edwin B. Parker Professor of Law, Columbia Law School, New York, USA; Professor, University of Neuchâtel, Switzerland. Email: pm2030@columbia.edu. I am grateful for many discussions on the issue with Cary Coglianese and Rodd Izadnia.

EC – Hormones, and *US – Tax Legislation (DISC) / US – FSC* disputes.[1] We only recently (December 2018) saw the end of it, or so it seems. In part at least, this is due to the fact that the courts of the World Trade Organization (WTO) had not managed to explain in clear terms what the USA should do. In turn, this is due to their convoluted understanding of the WTO Agreement on Technical Barriers to Trade (TBT Agreement).[2]

Following repeated condemnations of its measures by a GATT panel already in 1991, the United States eventually allowed all tuna, no matter how it had been harvested, to be sold in its market. It reserved nevertheless the label 'dolphin-safe' to tuna harvested in particular manner. Mexico complained and prevailed yet again. The United States modified some aspects of its measure, without however, touching upon its quintessential elements. The amount of proof required to show that tuna could legitimately have access to the 'dolphin-safe' label would depend on the likelihood that dolphin life had been accidentally taken when fishing for tuna (that is, the US measure was 'calibrated' to the likelihood of accidentally taking the life of dolphins when harvesting for tuna).

The most recent reincarnation of the US measure was finally found to be WTO-consistent by a 'compliance panel', that is, a WTO 'court' examining the consistency of US implementing measures with an adverse decision by a WTO panel or the Appellate Body. This report has already done a lot of good. It is not enough though, as it changed the outcome of the dispute without changing the test of consistency with the TBT that the Appellate Body had already elaborated.[3] This is what we attempt to do in this paper, since the test for consistency now applied, because of its endogenous incoherence, can lead to antithetical results (as this case demonstrates).

Our argument runs like this. The TBT Agreement is asking different questions than the GATT. Its function is not to insure against erosion of tariff concessions. It is that as well. Its main function though, is to ensure that non-tariff barriers (NTBs) will not be erected when they do not genuinely serve a legitimate regulatory objective. In this vein, the agreement imposes a three-prong test, whereby members will intervene only when necessary to advance an objective, will adopt the least restrictive measures when doing so and will apply adopted measures in a non-discriminatory manner. Furthermore, unilateral measures will be adopted, only if existing international standards, which are presumed to be necessary (for example, the least restrictive option to advance an objective), cannot be appropriately used.

1 Throughout this article, I refer to the official title of all GATT and WTO disputes, as they appear in the WTO webpage, where all reports are available (www.wto.org).

2 Agreement on Technical Barriers to Trade (TBT Agreement) 1994, 1868 UNTS 120.

3 The report was appealed and the Appellate Body issued its report on December 14, 2018, as the present paper was about to be published. The Appellate Body addressed only one of the issues that dealt with in this paper, namely whether the challenged measure was applied in an even-handed manner. The Appellate Body thus, did not have to re-think the legitimacy of the test of compliance with the TBT Agreement, as it had been applied in prior case law, in holistic manner. See the Appellate Body report on *US – Tuna II (Mexico) (Article 21.5-Mexico)* at pp. 25ff.

Thus, the architecture of the TBT Agreement inherently carries a sequence in the sense that members should first see to what extent an international standard can be appropriately used to serve their preferences. Recourse to unilateral measures is appropriate only when response to this question is negative. Consequently, WTO members are not free to adopt any measure they wish when regulating an issue coming under the aegis of the TBT Agreement. They have to use measures, in principle at least, that have been jointly negotiated in standard development organizations.

There is another sequence, though, embedded in the TBT Agreement. The key difference between the GATT and the TBT Agreement is that the latter requires the adoption not simply of non-discriminatory measures but also, crucially, of necessary measures (a subset of all possible non-discriminatory measures) that should be applied in an even-handed manner. Necessity is presumed when international standards have been privileged, but not when recourse to unilateral measures is made. This is the other restriction on regulatory discretion. In the GATT world, any non-discriminatory measure passes the test of consistency. In the TBT Agreement world, only a subset of non-discriminatory measures (the 'necessary' measures) does so. By requesting that necessary measures are applied in non-discriminatory manner, the TBT Agreement imposes a second sequence, namely, between the question whether a measure is necessary, and the question whether it has been applied in even-handed manner.

Crucially, the TBT Agreement, unlike Article III of the GATT, is not about market-like measures but, rather, about policy-like measures. Indeed, measures coming under the aegis of the TBT Agreement could concern different product markets. A WTO member could request, for example, that all goods circulating in its market carry a label indicating their environmental footprint.

To support our claim, we go through the case law on the TBT Agreement. We use the tuna disputes as background to explain what is wrong with the current test for consistency as developed by the Appellate Body and then advance our own preferred test for consistency with the TBT Agreement.

The errors by the Appellate Body in interpreting the TBT Agreement are not confined to the understanding of the obligations imposed on unilateral measures, such as those employed in the tuna disputes. The Appellate Body has created, unnecessarily so, havoc regarding the policy relevance of international standards, one of the pillars of the TBT Agreement. Peru had a narrow escape in *EC – Sardines*, when the Appellate Body forgot for a moment its 'textualist' self and adopted an interpretation of the term 'except', featured prominently in the body of Article 2.4 of the TBT Agreement, which is hard to reconcile with the intrinsic meaning of this term.

In short, case law under the TBT Agreement leaves a lot to be desired. Section 2 explains the test for consistency with the TBT Agreement as developed in successive Appellate Body reports. Section 3 is divided into two parts. We first discuss what, in our view, is still wrong with the test for consistency developed in the case law, the

recent improvements notwithstanding, and then we ask why this has been the case. It is in section 4 that we advance our suggested approach, while section 5 recaps the main conclusions.

2 The Current Test for Consistency with the TBT Agreement

A *TBT Agreement, TBT Disputes and International Commerce*

According to recent estimations, technical regulations and standards (the instruments coming under the aegis of the TBT Agreement) were potentially linked to 92 per cent of US goods exports in 2015, and 93 per cent of global exports in the same year.[4] Unsurprisingly, thus, the number of TBT-related disputes has been proliferating as well. The test of consistency with the TBT Agreement, when it comes to recourse to international standards, has been developed in *EC – Sardines* and in *US – Tuna II (Mexico)*. The test of consistency with the TBT Agreement, when it comes to unilateral acts, has been developed in three reports – namely, *US – Clove Cigarettes*, *US – COOL (Canada)* and *US – Tuna II (Mexico)*. The approach developed though, has had an impact on a much wider class of cases. For instance, there have been 52 disputes, where at least one claim has come under the ambit of the TBT Agreement, that have been formally raised since the advent of the WTO.[5] This suggests a rate of more than two cases yearly. To give an order of magnitude, this is, approximately, the annual workload of the International Court of Justice. In March 2016, the 500th specific trade concern (STC) was raised.[6] STCs are a hybrid between transparency and dispute adjudication.[7] One would expect that contract completion through the case law regarding the consistency with the TBT Agreement will affect the discussions in the realm of STCs. And, of course, the Agreement on Sanitary and Phyto-Sanitary Measures (SPS Agreement), which endorses a similar test for consistency and also its own STCs, will be heavily influenced as well by the test established under the TBT case law.[8] One can thus easily appreciate that a wide realm of WTO practice will be influenced – actually, is being influenced – by the current test for consistency.

4 J. Okun-Kozlowicki, *Standards and Regulations: Measuring the Link to Goods Trade*, Office of Standards and Investment Policy, Industry and Analysis, International Trade Administration, US Department of Commerce, Washington DC (2016).

5 Our cut-off date is 31 December 2017. All information regarding disputes formally raised before the World Trade Organization (WTO) can be accessed at www.wto.org.

6 WTO, 'DG Azevédo Praises Work of Standards Committee on Reaching 500th Trade Concern', 8–10 March 2016, available at www.wto.org/english/news_e/news16_e/tbt_11mar16_e.htm. To provide an order of magnitude, at the moment of writing (December 2018), 573 disputes have been formally raised before the WTO adjudicating bodies.

7 For a discussion on the origins, function and practice of specific trade concerns, see Horn, Mavroidis, and Wijkström, 'In the Shadow of the DSU: Addressing Specific Trade Concerns in the WTO SPS and TBT Committees', 47 *Journal of World Trade* (2013) 729.

8 Agreement on Sanitary and Phytosanitary Measures 1994, 1867 UNTS 493.

B *The Test of Consistency of Unilateral Acts with the TBT Agreement*

The Appellate Body has held that:

- the obligations regarding consistency of technical regulations (standards and conformity assessment) – namely, non-discrimination and necessity – are independent from each other.[9]
- Non-discrimination has two legs since the complainant must show that two goods are like and that imported goods are afforded less favourable treatment than domestic goods:
  - The term 'like products' should be understood in the same way as it is in the GATT case law – namely, that it is consumers that decide on likeness.
  - Conversely, the term 'less favourable treatment' should be understood in a GATT-unlike manner. Contrary to the GATT case law, regulatory distinctions that lead to disparate trade outcomes are not discriminatory if they are not origin based. Note, though, that the (disparate) trade outcome must be exclusively due to the (legitimate) regulatory distinction.
- Necessity has the same content in the GATT as in the TBT Agreement – that is, it requires from WTO members to reduce cost shifting by adopting the least trade restrictive measures in order to pursue legitimate regulatory objectives.[10]

1 Necessity

The leading case in TBT case law is *US – COOL*. There, the Appellate Body reproduced the classic GATT test for consistency, asking whether the labelling measure privileged by the USA had been the reasonably available least restrictive option to reach the stated objective. In *US – Tuna II (Mexico)*, the Appellate Body performed the necessity

[9] The TBT Agreement covers four instruments: international standards (e.g., standards adopted by the International Organization of Standardization), technical regulations, standards (instruments regulating the production process of goods compliance with which is necessary condition for market access with respect to the former, but not to the latter) and conformity assessment – that is, the procedure allowing the importing state to verify whether imports conform to the content of international standards, technical regulations or standards. International standards are presumed to meet the necessity criterion (Article 2.5 of the TBT Agreement). There is no case so far where a panel had to review a claim that an international standard had not met the necessity criterion. Although Articles 2.1 and 2.2 of the TBT Agreement do not formally apply to standards, Annex 3 (D, E, F and so on) leave us with no doubt that standards must be necessary and applied in a non-discriminatory manner. These obligations echo thus Articles 2.1 and 2.2 of the TBT Agreement. In a similar vein, Article 5 of the TBT Agreement imposes these obligations on conformity assessment procedures. For a detailed discussion of this issue, see P.C. Mavroidis, *The Regulation of International Trade* (2016), vol. 2, at 389ff.

[10] Of course, necessity is a positive obligation in the TBT Agreement and becomes an obligation in the GATT only if an exception has been invoked. What we mean here though is that the substantive content across the two provisions (Article 2.2 of the TBT Agreement; Article XX of the GATT) is the same in that it obliges the regulating state to reduce costs shifted to the international trading community by adopting the instrument that has the smallest impact on international trade. The findings of the three reports mentioned here are discussed in detail in Mavroidis, 'Driftin' Too Far from Shore', 12 *World Trade Review* (*WTR*) (2013) 509.

analysis under Article 2.1 of the TBT Agreement (non-discrimination), as we discuss in the following sections.

2 Non-Discrimination

In *US – Tuna II (Mexico)*, the Appellate Body asked whether consumers would treat dolphin-safe and dolphin-unsafe tuna as like products. Having responded affirmatively to this question, the Appellate Body went on to ask whether the regulatory distinction operated by the USA afforded less favourable treatment to Mexican tuna. It asked whether the (disparate) trade impact on Mexican producers was exclusively due to the legitimate regulatory distinction operated in US law. It went ahead to find that this measure was discriminatory because most Mexican vessels were fishing in the area with a high concentration of dolphins and, hence, had to observe the more onerous reporting requirements. The US measure, thus, was found to be discriminatory but necessary.

Paradoxically, in the same dispute, the Appellate Body performed under Article 2.1 of the TBT Agreement what seems to be a necessity analysis. Recall that the USA had in place a measure whereby vessels fishing for tuna observe enhanced verification and reporting requirements (certifying no incidental taking of dolphin life) when fishing in areas with a high concentration of dolphins than when fishing in areas with a low concentration of dolphins. The measure was thus calibrated to the likelihood that dolphins could be accidentally killed (high likelihood would lead to enhanced reporting requirements, whereas low likelihood would lead to 'lighter' requirements). The Appellate Body found that the US reporting requirements were effectively calibrated to the magnitude of risk of encountering dolphins when fishing for tuna. Thus, it was necessary to contribute towards achieving its stated objective. In doing that, it employed the classic GATT analysis regarding the 'necessity' requirement:

- In light of the objective pursued, is the measure the least restrictive option (that is, the option that has the least negative impact on international trade)?
- If yes, is it reasonably available to the regulating state?
- If no (if it imposes undue hardship on regulator), is the next in line least restrictive measure reasonably available to it?

3 Standards

In *US – Tuna II (Mexico)*, traders were not banned from accessing the US market if their tuna did not qualify as 'dolphin safe'. The US position was thus that the measure adopted was a standard and not a technical regulation. Mexico protested, arguing that the label conferred an advantage to those traders that could have access to it and that, by restricting the use of the 'dolphin-safe' label to only those fishers who used a specific harvesting technique, the USA had enacted a technical regulation, not a standard. The panel agreed with Mexico. What mattered was that traders did not have unconditional access to the use of the label.[11] A separate opinion issued by one member of the

[11] *US – Tuna II (Mexico)*, paras 7.100ff, especially paras 7.120, 7.131.

panel stressed that what was relevant for the qualification of the measure as a technical regulation or standard was ultimately whether imports of tuna could (or could not) be lawfully traded in the US market without the label 'dolphin safe'. Evidence showed that Mexican traders, whose products did not conform to the criteria established in US law, could still sell their products in the US market if they marketed them as 'tuna', not 'dolphin-safe tuna'. Consequently, the measure at hand, in this view, was a standard.[12]

The Appellate Body endorsed the majority view, holding that the US measure was indeed a technical regulation.[13] In its view, what mattered was that goods had to observe the statutory requirements (and had no discretion at all to this effect), otherwise, they could not legitimately be considered 'dolphin safe'.

C *The Test of Consistency with International Standards (Burden of Proof)*

In *EC – Sardines*, the Appellate Body held that, in case of deviation from an international standard, the allocation of the burden of production of proof should not shift to the member deviating from the standard (and, obviously, possessing the information why it had done so).[14] The Appellate Body justified its approach on two distinct grounds.

- It held that deviation from international standards is not an exceptional circumstance per se.
- Furthermore, the Appellate Body also stated that the complainant, who assumes the burden of proof, would not be disadvantaged anyway since the TBT Agreement contained many transparency obligations. As a result, all WTO members are well acquainted with measures coming under its ambit and are thus in a position to argue why, in their view, notified measures have violated the TBT Agreement.

Implicitly, in the Appellate Body's view at least, we are not facing a state of asymmetric information that would have justified a different allocation of the burden of proof, precisely because of the elaborate transparency obligations embedded in the TBT Agreement.

D *The Facts before the Second Compliance Panel*

Following the condemnation of its policies, the USA adopted implementing legislation in order to address the concerns of the Appellate Body. Its first effort to implement was judged to be inconsistent with its obligations. The USA made a second attempt to implement, and Mexico once again disagreed that this had been the case and complained to the WTO yet again. In *US – Tuna II (Mexico) (Article 21.5 Second Recourse)*,

[12] *Ibid.*, paras 7.146ff.

[13] *Ibid.*, paras 172–199.

[14] *EC – Sardines*, paras 269ff. The repercussions of this approach to the (continuing) relevance of international standards in WTO law have been elaborated in Horn and Weiler, 'European Communities – Measures Affecting Asbestos and Asbestos-Containing Products', 3 *WTR* (2004) 129.

the second implementing effort by the USA, the US measure, the consistency of which with the WTO rules was at stake,[15] was as follows:

- Tuna caught through two fishing methods – namely, purse seine nets with setting on dolphins and driftnet – can never be eligible for the 'dolphin-safe' label.
- Tuna caught using other methods of fishing are, in principle, eligible for the 'dolphin-safe' label if it is proved that no dolphins were killed or were seriously injured through the gear deployment or fishing method used (these are the so-called 'eligibility' criteria, in the sense that tuna caught with any but the two aforementioned methods are, in principle, eligible to access the 'dolphin-safe' label).
- To prove that tuna can legitimately carry the 'dolphin-safe' label, interested producers must observe certification, tracking and verification, as well as determination, requirements:
  - For tuna caught in the Eastern Tropical Pacific (ETP), certification requirements are heavier: both the captain of the ship as well as an independent observer must attest to the fact that dolphins were killed or injured.
  - For tuna caught outside the ETP, certification requirements are 'lighter': the captain of the boat, having received training to this effect, can confirm whether dolphins were killed or injured, and there is no need to also have an observer on board.
  - The differentiation between ETP and non-ETP tuna is predicated on the fact that it is in the ETP that a high concentration of dolphins is observed and most of the setting on of dolphins takes place.
  - The US authorities can determine that an observer must be on board and certify that dolphin life has not been endangered even for non-ETP fisheries, if the fisher(s) at hand have a record of systematic accidental killing or endangerment of dolphin life.
  - Finally, the new, modified measure added heightened tracking and verification requirements (processors and importers of tuna must collect and keep information for two years so as to enable the US authorities to track non-'dolphin-safe' tuna back to the timing of its harvesting). Not only fishers harvesting tuna in the ETP, but also fishers outside the ETP, have to observe this requirement as well.

The US measure is thus calibrated to an 'exogenous' event – namely, the likelihood to endanger dolphin life when fishing for tuna in areas with high and low concentrations of dolphins. It is thus markedly unlike the situation in *EU – Seal Products*, where the

[15] A side remark is in order here. It is questionable whether the now well-embedded practice of having two compliance panels in the same dispute is consistent with the letter and the spirit of the WTO Understanding on Rules and Procedures Governing the Settlement of Disputes 1994, 1869 UNTS 401. Losing defendants in WTO litigation have only one reasonable period of time within which they can bring their measures into compliance. The second compliance panel by definition examines measures that were adopted outside the reasonable period of time – that is, new measures – which should normally be adjudicated before a new panel. As said though, all of this is by now water under the bridge.

challenged measure had been 'calibrated' to whatever the domestic political economy could take.

1 What Changed with This Report?

The panel had in front of it one important new element only: the new tracking and verification requirements that the US authorities were now imposing on fisheries outside the ETP. Calibration of the measure to the likelihood of endangering dolphin life continued to be the key element of the US measure. The objectives of the US measure were the same – namely, the protection of consumers as well as the protection of dolphin life, as the panel itself explicitly acknowledged.[16] The regulatory changes between the first and the second compliance procedure should thus not be exaggerated.

To decide whether the US measure was discriminatory, the panel focused its attention on whether it afforded less favourable treatment to Mexican tuna. It decided that this had not been the case.[17] It first echoed the legal standard developed by the Appellate Body, according to which no less favourable treatment can result if disparate effects for imported goods are exclusively attributed to a legitimate regulatory distinction. In line with previous panels and the Appellate Body, it construed that the TBT Agreement *à la* the GATT was an instrument protecting competitive conditions. Consequently, this panel did not look into actual trade effects. It held that, because the variation in the regulatory requirements had been calibrated to the risk, no less favourable treatment could have ever resulted for Mexican tuna. The Appellate Body confirmed.

2 What Stayed the Same?

This panel, and consequently, the Appellate Body as well, did not touch on the understanding of the term 'like products'. The test for likeness stays in the marketplace, and it is consumers based on the physical characteristics, properties and end uses of the product that will decide whether two goods are like or not.[18]

3 End Result: USA Wins

Against this background, the panel found that the US measure was not discriminatory and exonerated the USA from responsibility, and the Appellate Body confirmed. The outcome makes sense (at least to this author), but the reasoning still leaves a lot to be desired.[19] We take particular issue with the understanding of 'like products' (as it had been developed in the original panel, as well as in all other panels dealing with claims under the TBT Agreement), and the absence of sequence (analysis of Article 2.2 of the TBT Agreement before Article 2.1) for the reasons that we now develop.

[16] *US – Tuna II (Mexico)* (Article 21.5 Second Recourse), para. 7.49.

[17] *Ibid.*, paras 7.529–717.

[18] *Ibid.*, para. 7.74.

[19] Recall that it was the introduction of flexibility (in the sense of allowing for trading of shrimps harvested through fishing methods other than turtle-excluding devices) that was the lynchpin for the Appellate Body to accept that the USA had brought its measure into compliance with its obligations in the *US – Shrimp* litigation. And it is introduction of flexibility that did the trick in the tuna dispute as well.

3 What Is Wrong with the Current Understanding of the TBT Agreement and Why?

The Appellate Body in its case law so far has committed, in our view, four errors when addressing claims under the TBT:

- It interpreted the two obligations regarding consistency of technical regulations and/or standards with the TBT Agreement (namely, non-discrimination and necessity) as if they were independent from each other.
- Because of this error, it has also adopted an erroneous understanding of the term 'like products'.
- It reduced the relevance of international standards, against the expressed negotiating will.
- Finally, the distinction between 'technical regulations', on the one hand, and 'standards', on the other, became a line in the sand (although statutory language had set it in stone).

The recent panel report on *US – Tuna II (Mexico)* (Article 21.5 Second Recourse) has addressed none of them. Consequently, our criticism of the current case law stands. In what follows, we first explain why, in our view, the four interpretations are erroneous. We then argue that all of these errors are due to the fact that the Appellate Body has failed to see the objective function of the TBT Agreement as distinct from that of the GATT.

A *Why Are These Four Interpretations Erroneous?*

One of the main claims in this article is that the current test for consistency with unilateral measures is erroneous because the Appellate Body has not sequenced its analysis of the obligation reflected in Article 2.1 (non-discrimination) to that included in Article 2.2 (necessity) of the TBT Agreement. It has understood them as parallel obligations. In our view, this is wrong for the reasons explained below. The whole logic of the TBT Agreement is to impose a restraint on regulatory activity, as we explain in more detail in section 3.B. It does so by sequencing unilateral action to the inefficacy or inappropriateness (or even inexistence) of international standards. It does so also by asking regulators to think of the necessity to intervene at all. And it does so finally by imposing an obligation to adopt the least restrictive measure when regulating. Non-discrimination is meant to ensure that the otherwise necessary measures have not been implemented in a manner that does not observe an even-handedness requirement. The fact that Article 2.2 of the TBT Agreement (necessity) follows Article 2.1 (non-discrimination) is not persuasive evidence to the opposite. Indeed, it is the same Appellate Body that in its *US – Shrimp* jurisprudence sequenced its analysis of the chapeau of Article XX of the GATT (which echoes *mutatis mutandis* Article 2.1 of the TBT Agreement) to that of an evaluation of consistency with a sub-paragraph of the same provision (which echoes almost verbatim Article 2.2 of the TBT Agreement).

1 *Lack of Sequence Can Lead to Absurd Results*

Assume that the tuna dispute had been adjudicated under the GATT and not under the TBT Agreement. Article III.4 of the GATT would have been the relevant provision to entertain the dispute. The panel would have to ask if 'dolphin-safe' and 'dolphin-unsafe' tuna were like products. Since consumers define likeness, as we know from consistent case law, the response to this question should be in the affirmative.[20] The next question then would be whether the USA, by allowing (some) US tuna to carry the 'dolphin-safe' label and disallowing Mexican tuna to do the same, would be according the latter less favourable treatment than the former. GATT case law has gone full circle here. In *Korea – Various Measures on Beef*, the Appellate Body had held that the modification of conditions of competition to the detriment of imported goods equalled less favourable treatment.[21] Later, in *Dominican Republic – Import and Sale of Cigarettes*, the Appellate Body held that detrimental effects would not amount to less favourable treatment if the reason for it was a regulatory distinction unrelated to the origin of goods traded.[22] In *European Communities (EC) – Seal Products*, the Appellate Body returned to its Korea understanding and, thus, held that all disparate effects would amount to less favourable treatment.[23]

If the Appellate Body followed its jurisprudence under the *Dominican Republic – Import and Sale of Cigarettes* case, then it would have concluded that no violation of Article III.4 of the GATT had occurred. If it followed its jurisprudence under either of the other two disputes, the opposite would have been the case. Defendants would then have to invoke Article XX of the GATT to justify their violation. This is what the USA, in our case, would do. Assume that the USA chose to invoke Article XX(b) of the GATT.[24] The measure would have been considered necessary. Indeed, labelling is the most innocuous of the options available to the USA. In this case, labelling is based on an analysis regarding calibration to risk that the Appellate Body of which has approved. The only remaining question would be whether the US measure had been applied in an even-handed manner.

Recall *US – Shrimp*. Once persuaded that the measure (banning imports of shrimps that had not been harvested with turtle-excluding devices [TEDs]) was relating to the protection of an exhaustible natural resource (sea turtles), the Appellate Body asked

[20] Assuming, of course, that all consumers valued dolphin life enough and were prepared to always pay a higher price for dolphin-safe tuna, then the result could be the opposite. This is a rather unrealistic assumption.

[21] *Korea – Various Measures on Beef*, para. 149. In Mavroidis, *supra* note 9, at 385ff, we held that this is a deplorable outcome. The Korean measure had led to fewer outlets selling imported beef because of the legal import quota that had been in place (since traders would rather sell the cheaper imported beef rather than the more expensive domestic beef). Furthermore, all of the quotas for imported beef had been routinely sold in the Korean market. If at all, the number of outlets should have been treated as a service, not a goods, issue.

[22] *Dominican Republic – Import and Sale of Cigarettes*, para. 96.

[23] *European Communities (EC) – Seal Products*, paras. 5.109–110.

[24] Following *US – Shrimp*, it could, of course, invoke Article XX(g) of the GATT since living organisms as well are considered 'exhaustible natural resources' and, thus, have an even easier road to victory. It does not matter though, for the reasons I explain here.

whether the measure had been applied in a non-discriminatory (even-handed) manner. It held that it was not the case. The reasoning nevertheless matters. It concluded as much only because the US objective could have been reached by allowing imports of shrimp that had been harvested through methods other than, but as efficient as the use of, TEDs. Implicitly, thus, it accepted the legitimacy of the US distinction between shrimp harvested with TEDs (as well as with other equally efficient methods) and shrimp harvested in a manner that endangered the life of sea turtles. It did not ask consumers whether they would treat the two types of shrimp as like goods. Thus, the Appellate Body endorsed that, in the realm of Article XX of the GATT, there is no more room for market likeness. Indeed, it is regulatory intent that becomes the key in deciding on the compliance of measures that a WTO member seeks to justify through recourse to Article XX of the GATT.

Think now of ETP and non-ETP as two WTO members. In the former, the risk of killing dolphins is high; in the latter, it is not. These are not two countries where the same conditions prevail. Thus, imposing differentiated regulatory requirements would be a perfectly legitimate measure in light of the regulatory objective pursued. Were the USA to treat all tuna harvested in the ETP under one demanding set of requirements and all tuna harvested in the non-ETP under another, less demanding set of requirements, it would be adhering to the even-handedness requirement embedded in the chapeau of Article XX of the GATT, as per *US – Shrimp*.

If the USA would have prevailed under Article XX(b) of the GATT, as we have shown here, should it then be on the receiving end under the TBT Agreement? This is a question the (original) Appellate Body should have asked, but it did not ask. The US measure was found twice to violate the TBT Agreement, even though it would have sailed through the GATT. What is even more surprising is that the USA lost when the allocation of the burden of proof was more favourable to it than it was in the GATT world (where it would have prevailed). Under the TBT Agreement, the burden of proof stays with Mexico, whereas in the GATT context it shifts to the USA only after Mexico has established a violation of Article III of the GATT.

The TBT Agreement, of course, adds to the GATT, otherwise why sign the TBT Agreement in the first place? There is an obligation to start from implementing international standards, which do not exist anywhere in the GATT. This, nevertheless, was a non-issue here since the standard invoked by Mexico was not considered already to be an international standard in the original Appellate Body report.[25]

2 Consumers' Preferences Are Immaterial

In the TBT context, consumers' perceptions for likeness are immaterial. To start with, if government and consumers saw eye to eye – if they shared preferences – there would be no need to enact technical regulations or standards. Governments intervene either

[25] Mexico had claimed that the Agreement on International Dolphin Conservation Program (AIDCP) standard was international in the TBT sense of the term. The Appellate Body held that this was not the case as the AIDCP was not open to all WTO members. *US – Tuna II (Mexico)*, paras 343ff.

because they have private information (this is very often the case, for example, in the realm of environmental legislation) or simply because they have different preferences from (some) consumers (as in our case). Governments, of course, have the right to do that under domestic constitutions, and the WTO (TBT Agreement) has not questioned this right. After all, the WTO is a negative integration contract, where preferences are set unilaterally and, to the extent that they shift costs to trading partners, a WTO member must observe non-discrimination (and, in the case of the TBT Agreement, the necessity principle as well).

What matters thus, is not the consumer definition of likeness. What matters is the government definition of likeness. Likeness in the TBT world is thus not market likeness. It is policy likeness. It is for the USA to decide on what is 'dolphin-safe' tuna. What if the USA, when doing so, confers a regulatory subsidy to its domestic producers? Well, the TBT Agreement contains, of course, important safeguards against abusive interventions. First, assuming no international standard exists, governments must think whether interventions are necessary at all. They should go ahead only in presence of evidence that non-intervention would be costly. Second, measures must be necessary to achieve a legitimate regulatory objective. The agreement contains an indicative list to this effect, which informs users of what the framers had in mind and helps adjudicators to avoid false negatives. Third, and assuming its measure is necessary – that is, assuming the 'dolphin-safe' label represents the least restrictive means to reach its ends (protection of consumers and life of dolphins) – the measure must be applied in a non-discriminatory manner.

Thus, the only remaining question for the Appellate Body, having satisfied itself on the necessity of the US measure, should be whether Mexican tuna that meets the US definition of 'dolphin safe' has been treated in the same manner as US 'dolphin-safe' tuna. Why is this case? Well, by accepting the necessity to distinguish between 'dolphin-safe' and 'dolphin-unsafe' tuna, the Appellate Body has accepted the legitimacy of distinguishing between two classes of like products (in the eyes of the consumers) in order to serve the overall legitimate objective – namely, the protection of consumers and the protection of dolphin life. The only remaining question should be whether goods falling under each category are treated in an even-handed manner, irrespective of their origin.

Going back to market likeness is tantamount to questioning the right of the USA to intervene in the first place and distinguish between tuna fished in one or the other way. This is precisely the question the TBT Agreement asks through the requirement of necessity, which is a double requirement, to be sure: first, is it necessary for the USA to intervene? Second, if yes, has the USA adopted the necessary (least restrictive) measure to this effect? If yes, then all the USA has to comply with is a requirement of even-handedness. Think of a case where an international standard has been used. Assume for example, that an International Organization of Standardization (ISO) standard exists for 'dolphin-safe' tuna.[26] Assume further, that the USA refuses to

[26] It is the only standard development organization explicitly mentioned in the TBT Agreement.

apply it in non-discriminatory manner, the claims by Mexicans that their tuna meets the standard notwithstanding. Would the Appellate Body in this case ask what the US consumers' view on likeness is? The short answer is no. International standards are presumed necessary (Article 2.5) but must be applied in a non-discriminatory manner. Technical regulations and/or standards are not presumed necessary. The inquiry into their necessity is meant to give the green light to a government to impose a certain behaviour on its citizens. To the extent that technical regulations and/or standards have been found to comply with the necessity requirement, it is their content, and not consumers' preferences, that should provide the benchmark for likeness.

Measures coming under the TBT Agreement often cut across various distinct product markets. The marketplace test to decide on likeness of goods in the GATT is predicated on the existence of a precisely defined market. Think of labelling the environmental footprint of all goods circulating in a WTO member. In similar cases, the question will not be whether, say, shoes produced with or without 'green' technology are like. The question here is whether labelling of the environmental footprint is necessary to achieve the regulatory objective sought. If yes, then the question will arise whether foreign goods meeting this government-imposed standard meet the requirements embedded in the labelling scheme. Consumers' preferences are simply immaterial in this exercise.

There is, of course, room for necessary measures to be found discriminatory, in case imported goods are conforming to the substantive requirements that the regulating state has enacted, but still do not profit from the same treatment as domestic products conforming to the same requirements. Conformity assessment holds the key in distinguishing the wheat from the chaff.

3 Need to Dis-Entangle Necessity from Likeness Analysis

Recall our discussion above concerning the treatment of calibration of risk under the heading 'non-discrimination'. The Appellate Body should have performed this analysis under Article 2.2 of the TBT Agreement and not under Article 2.1 of the TBT Agreement. Calibration to the risk is the quintessential element of the challenged US legislation. All reporting requirements, and, indeed, the labelling scheme itself, are based on the calibration of the US measure to the risk of accidentally killing dolphins. It is impossible to pronounce on the necessity of the measure, unless the calibration to risk has been taken on board.

In *US – Tuna II (Mexico)*, the Appellate Body conflated the two issues in one and did not even enter the discussion of necessity of the measure. How can it be though that it is consumers that decide on likeness, but calibration to risk, a pure regulatory measure, forms an integral part of the non-discrimination analysis? Where is the proof that US consumers were privy to the information that the dolphin-safe measure had been calibrated to the likelihood of accidentally taking the life of dolphins when harvesting tuna? If the Appellate Body uses its calibration analysis as part of its inquiry into whether the USA is affording imported goods less favourable treatment, then it is making yet another mistake. It is effectively asking if the USA had the right

to intervene in the first place, instead of asking whether the measure has been applied in a non-discriminatory manner, as it should, when examining the consistency of any measure with Article 2.1 of the TBT Agreement. By dis-entangling, as suggested above, non-discrimination has an autonomous existence. WTO courts will be in a position to examine claims pertaining solely to Article 2.1 of the TBT Agreement, assuming the necessity of the measure has not been challenged.

4 Detriment Exclusively Caused by Legitimate Distinction

The Appellate Body, in its report on *US – Tuna II (Mexico)*, held that a measure producing detrimental effects will still not be considered to accord less favourable treatment to imported goods if the detrimental impact is due exclusively to a legitimate regulatory distinction.[27] This is a very demanding test and quite unwarranted if the purpose of this provision, as the Appellate Body itself has explicitly admitted in numerous reports, is to ensure that domestic and imported goods will be treated in an even-handed manner. We explain.

If we were to unravel the test as applied, we would first need to be clear about its constituent elements:

- if the challenged measure modifies the conditions of competition;
- to the detriment of imported goods;
- and the detrimental impact results exclusively from a legitimate regulatory distinction;
- then, the measure does not afford less favourable treatment to imported goods.

The modification of conditions of competition should not be an issue, since WTO members retain the right to adopt any policies they wish to the extent that they apply them in an even-handed manner.[28] The first question is how we should measure the relative detrimental impact? Imported goods must be hit harder than domestic goods, but by how much? The simple answer is that we do not know. The Appellate Body has discarded the relevance of trade effects. So, the best evaluation we can have here is some sort of sophisticated qualitative guess as to how trade trends will evolve because of the measure adopted (and challenged).

But then comes the even bigger thorn. Assuming the Appellate Body guesstimates that the hit will be harder on imported goods, it must then ask whether the (contemplated) impact results exclusively from a legitimate regulatory distinction. This is quite a hurdle because the TBT Agreement does not request the adoption of first best but, rather, of necessary policies. Negative external effects are minimized, even though there is no absolute guarantee that they are eliminated altogether, when first-best policies are adopted.[29] When necessary policies are adopted, then we are by definition in the

[27] *US – Tuna II (Mexico)*, para. 297.

[28] Modification of the conditions of competition that do not result in detrimental impact for imported goods can still be the object of a non-violation complaint. The study of this eventuality though escapes the purview of this article.

[29] The foundational paper in the targeting literature is Bhagwati and Ramaswami, 'Domestic Distortions, Tariffs, and the Theory of Optimum Subsidy', 71 *Journal of Political Economy* (1963) 44.

presence of some burden on international trade. Furthermore, in a realistic scenario, other factors (other than regulation) might influence the outcome as well. Is it, for example, crystal clear that Mexican producers suffered damage exclusively from the 'dolphin-safe' label? What if some consumers would anyway buy 'dolphin-safe' tuna? Or what if fish lovers turn to halibut because there has been an oversupply of halibut in a given year in the US market, and, as a result, its price is now much lower than that of tuna?

The Appellate Body routinely turns a blind eye to any sophisticated analysis when various factors affect one outcome, and this is irrespective whether we are dealing with a dispute in the TBT Agreement, the Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade, or the Agreement on Subsidies and Countervailing Measures context.[30] As a result, panels might be tempted to effectively pay lip service only to this requirement in future practice. Indeed, the recent panel report on *US – Tuna II (Mexico) (Article 21.5 Mexico)* already has found that the US measure met the exclusivity requirement because it was calibrated to the likelihood of endangering the life of dolphins.[31] This finding might make sense when it comes to discussing the necessity requirement. We need more information though to decide whether this is the case when it comes to answering the question whether the detrimental impact was due exclusively to the legitimate regulatory distinction.

5 *Allocation of Burden of Proof Is Wrong*

In *EC – Sardines*, Peru argued before the panel that, since it was the European Union (EU) that had deviated from an appropriate international standard, it was for it to explain why. The panel agreed.[32] The Appellate Body overturned this finding, arguing that there was nothing exceptional about a deviation from an international standard, adding that Peru, because of the elaborate transparency-related obligations embedded in the TBT Agreement, should be well aware of the rationale for the measures adopted by the EU.[33]

There are good arguments why this allocation of the burden of proof should be discarded in future practice. First, there is a clear legislative mandate stating that international standards must be followed, except when they cannot appropriately serve

[30] Agreement on Subsidies and Countervailing Measures 1994, 1869 UNTS 14; Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994, 1868 UNTS 201. Indeed, in the very recent *EU –PET (Pakistan)*, the Appellate Body reaffirmed its prior position to the effect that the methodology for separating effects caused by subsidies and other factors is not prejudged (hence, unsophisticated non-econometric analysis can, in principle fit the bill, and all that matters is that effects from subsidies are not marginal, without defining what marginal is, see paras 5.169ff, especially 5.175).

[31] *US – Tuna II (Mexico)* (Article 21.5 Mexico), para. 7.717. The Appellate Body confirmed. We read in para. 7.2 of its report: "[I]n order to determine whether the detrimental impact stems exclusively from a legitimate regulatory distinction, a panel must carefully scrutinize whether the technical regulation at issue is even-handed in light of the particular circumstances of the case." This passage is a far cry from any analysis aiming to show whether an independent variable is the exclusive cause of the dependent variable in question.

[32] *EC – Sardines*, paras 7.48ff.

[33] *Ibid.*, paras 285ff. The Appellate Body has repeated this allocation of burden of proof in subsequent cases and most recently in its report on *Indonesia – Import Licensing Regimes (United States)*, paras 5.49ff.

as a basis. The term 'except' can only be understood as referring to an 'exception'. In all reports dealing with claims to Article XX of the GATT, the Appellate Body has explained that, because this provision is an exception to obligations assumed, the burden of proof must shift to the party invoking one of the grounds mentioned therein. Second, what if the EU had withheld information from its partners? What if it had behaved in a non-transparent manner? Indeed, in cases of voluntary deviation from an international standard, similar behaviour would be incentive compatible. How then can the transparency obligations be of help to complainants in similar circumstances? Could Peru possess the rationale for the EU's decision to deviate from the international standard, had the latter decided not to communicate it to the TBT Committee?

Or what if the notification is inadequate? What if, for example, the notifying member failed to explain in detail the rationale for a measure that deviates from an international standard? Cary Coglianese usefully distinguishes between 'fishbowl' and 'reasoned' transparency.[34] The first covers cases where a measure is simply described. The latter extends to explanations of the rationale for the notified measure. According to the TBT Agreement (Articles 2.9.1 and 2.9.2), a member must submit a brief description of the objective and rationale of the measure, and additional information will be provided only upon request (Article 2.9.3). How then can complainants protect their interests when they have only partial knowledge of the objectives and rationale of the challenged measure? Did the Appellate Body implicitly add a requirement of due diligence, which would require WTO members to ask questions regarding the rationale for the measure (as opposed to enjoying discretion to this effect, as per the current architecture of the TBT Agreement)? Recall that it is the same Appellate Body that has consistently held that the policy rationale matters when it comes to whether the non-discrimination obligation has been adhered to.

One might further ask from a policy perspective whether this finding eviscerates the incentive to adopt international standards. Indeed, harmonization is one of the key objectives of the TBT, and relegating the policy relevance of international standards might act as a disincentive to pursue harmonizing efforts (and, thus, defeat a key objective of the TBT Agreement).[35] To complete the discussion, we should probably mention that Peru managed to win its argument. Nevertheless, it is quite remarkable that Peru did not add anything before the Appellate Body to what it had already argued before the panel. That is, Peru won by providing the same quantum of evidence in a context where the burden of the production of proof rested with the other party and in a context where it had to assume the burden of proof itself. This is, in and of itself, quite telling of the functionality of the allocation of the burden of the production of proof operated by the Appellate Body.

[34] Coglianese, 'The Transparency President? The Obama Administration and Open Government', 22 *Governance* (2009) 529.

[35] On this score, see the excellent analysis in Horn and Weiler, *supra* note 14. The authors underscore that by lowering the burden of persuasion of Peru, the relevance of international standards has been strengthened (at 138ff). This sounds plausible. The question, though, remains whether panels and the Appellate Body will consistently be satisfied with a similar burden of persuasion in future case law as well. This remains to be seen. What is clear for now is that the burden of production of proof does not switch in the case of deviation from international standard.

6 No Room for Standards Anymore?

Standards, as opposed to technical regulations, do not condition market access upon the satisfaction of statutory requirements. They can and do have a market effect, but they do not exhibit the binary function of technical regulations: either products meet the statutory requirements and are granted market access or they do not and stay out of the market as a result. In the case before us, were the USA to accept only 'dolphin-safe' tuna in its market, then, yes, we would have been in the presence of a technical regulation. By allowing 'dolphin-safe' and 'dolphin-unsafe' tuna into its market, the USA thought (legitimately so) that it had enacted a standard.

The Appellate Body disagreed. It held that, because the USA was not allowing all traders to use the 'dolphin-safe' label, it was effectively (de facto) imposing a technical regulation on Mexican fishers. It seems that the Appellate Body confused two separate issues: the distinction between technical regulations and standards, on the one hand, and a conformity assessment, on the other. Consistency with standards requirements must be safeguarded, of course. Standards are denied their *raison d'être*, if recourse to them is open even for products that fall short of meeting the established statutory requirements. Consumers will not be protected at all if, for example, products can carry the 'dolphin-safe' label regardless of whether the tuna sold has been fished in a dolphin-friendly manner. Recall that all panels and the Appellate Body acknowledged that the objective pursued by the USA was to inform consumers about the manner in which traded tuna had been fished. Opening up the use to the 'dolphin-safe' label to all tuna marketed, regardless of the harvesting method, would have led to misinforming consumers. The USA would never have achieved its stated (legitimate) objective. By requiring a conformity assessment though, the USA does not condition access to its market only to dolphin-safe tuna. It simply ensures that consumers will not be misled.

B *What Drives These Erroneous Interpretations?*

The Appellate Body has committed the cardinal mistake that no interpreter should ever commit. It did not ask the question what is the rationale explaining the advent of the TBT Agreement? It transposed, in a haphazard manner as we have noted, its interpretation of terms that the GATT and the TBT Agreement share. The same terms, nevertheless, can have a different meaning in two different contexts. One does not have to delve into Ludwig Wittgenstein to reject the view that words are 'invariances' – for example, context independent.

1 GATT-Think

GATT-think is genial in its simplicity.[36] 'Protection', an elusive concept,[37] is reduced to one instrument – namely, tariffs. Disciplines on all other instruments affecting trade

[36] The full negotiating record on this score has been presented in D.A. Irwin, P.C. Mavroidis and A.O. Sykes, *The Genesis of the GATT* (2008).

[37] Protection cannot be judged by effects of a measure, since, in equilibrium, any measure will have (at least indirect, potential) effect on the market. Regulator has private information, with no incentive to share it, since it might be self-incriminating. We are thus, facing a classic prisoner's dilemma.

(including domestic instruments such as those now covered by the TBT Agreement) aim at ensuring that the value of concessions will be eroded. This insurance policy has been necessary, otherwise trading partners might have lost the incentive to continue negotiating the level of tariffs downwards.

Robert Baldwin, in his classic account, explained the rationale behind the GATT integration process.[38] Tariffs and import quotas were blurring the size of the bite of 'snags' that were lying behind the border. They had to be tamed first. The 'snags', of course, were the various domestic policies. Because of policy substitution, domestic instruments could play the role of tariffs. The rational short- to mid-term solution was to make sure that snags do not bite imported goods more than their domestic counterparts with which they were in competition until they could be properly evaluated. Non-discrimination would reduce the impact of the bite. The consequence was that costs would be shifted in an even-handed manner to domestic and foreign products.

Unnecessary costs, nonetheless, could be shifted anyway since trading nations do not have to 'efficiently' address distortions or even address distortions at all. Any regulation, to the extent non-discriminatory, would pass the test of consistency with the GATT, even if it results in unnecessary costs for international trade and even if the instrument used to achieve an objective is totally inefficient.[39] At the heart of Article III is a common understanding that protectionist policies – that is, plain vanilla policies aiming to favour the domestic producer – should not substitute for customs duties (which do the same). Indeed, even in the WTO era, some of the disputes adjudicated under Article III concerned cases where tax differentials were not meant to serve a legitimate objective. They concerned tax differentials in favour of products produced predominantly by domestic producers. *Japan – Alcoholic Beverages II* is a very appropriate illustration.

Under the circumstances, it is only normal that panels established a test where interventions would be judged first by checking consumers' reactions (market likeness) and then, assuming government and consumers did not share preferences in this respect, by asking the question whether less favourable treatment had been afforded to imported goods as a result. This approach does not bode with TBT-think for the reasons expressed in the section that follows.

2 TBT-Think

In a world where NTBs essentially fragment markets, the content of domestic regulatory policies matters. Terms like harmonization, (mutual) recognition and a combination of the two have entered the vocabulary of the world trade lexicon. The TBT Agreement was

[38] R.E. Baldwin, *Non-Tariff Distortions in International Trade* (1970).

[39] One might retort, why would trading nations shoot themselves on the foot? Political economy, as well as the presence of prospective remedies in the GATT/WTO legal order, might help explain why this can be the case. Grossman, Horn and Mavroidis, 'Domestic Instruments', in H. Horn and P.C. Mavroidis (eds), *Legal and Economic Principles of World Trade Law*, 205; American Law Institute, *Reporters' Studies on WTO Law* (2013) discuss this point in depth.

the 'vehicle' that helped introduce this terminology into the GATT first and into the WTO later. Why the TBT?[40] The drastic reduction of tariffs entailed that domestic lobbies had to look elsewhere for protection. The (almost) irrelevance of tariffs, as well as the enhanced disciplining of subsidies, entailed that regulatory subsidies emerged as the obvious candidate for those seeking protection. If they succumbed to similar demands for domestic political economy reasons, and adopted legislation that copied domestic production patterns, trading nations would be shifting costs to foreign producers.[41]

The framers of the TBT Agreement laid out a legal discipline for judges to follow. Harmonize when necessary, recognize when possible, and anyway adopt necessary measures – this is the TBT Agreement recipe.[42] The TBT Agreement does not go as far as to impose common policies. It is not an instrument mandating positive integration. It is an instrument that seeks to minimize shifting costs to trading partners. This was a concern only with respect to border instruments (tariffs; import quotas) in the GATT-think. This is the concern for domestic instruments in the TBT world. International standards are presumed necessary. When no international standards exist (or when they cannot be appropriately used), WTO members should first think of the necessity to intervene, and, if the response to this question is affirmative, they should adopt necessary measures (Article 2.2). Necessity thus emerges as the linchpin of the whole edifice, since it must be observed irrespective whether recourse to international standards or unilateral measures is privileged. Necessity must be respected at all times.

Non-discrimination, in the TBT world, concerns only the application of necessary measures (either international standards or domestic technical regulations, standards and/or conformity assessment). Armed with this understanding of the TBT Agreement, we now propose an approach that WTO courts should adopt when adjudicating disputes coming under the aegis of the TBT Agreement.

4 Suggested Approach

No legislative change is required for our suggested approach to be implemented. In fact, we believe that the approach advocated here is the only one that is faithful to the current drafting of the TBT Agreement.

[40] The negotiating record of the first TBT Agreement is discussed in L.A. Brien, *The Tokyo Round Agreements, Technical Barriers to Trade* (1981), vol. 4; G.R. Winham, *International Trade and the Tokyo Round Negotiations* (1986); A.O. Sykes, *Product Standards for Internationally Integrated Goods Markets* (1995). They discuss the rationale for the second TBT Agreement, which is currently in force.

[41] T. Büthe and W. Mattli, *The New Global Rulers* (2011), at 85ff, mention some hilarious examples.

[42] Necessary means not the absolutely least restrictive measure, but the least restrictive measure available to the regulating WTO member. This is a case law contribution that the Appellate Body first introduced in its report on *US – Gambling*. The rationale, which we applaud, is that, otherwise, regulating states might have found it impossible to pursue legitimate objectives simply because pursuing them could be quite costly. The 'hardship test' that the Appellate Body has introduced aims at allowing members to use the next in line necessary measure, which they can employ without incurring a disproportionate cost.

A *Sequence Matters*

Assuming no international standard exists, or can be appropriately used, WTO courts should first ask whether a measure is necessary in the sense that it is the least restrictive option reasonably available to the regulating WTO member in order to reach its unilaterally defined objective:

- If the response is no, the WTO courts would not need to look any further. Violation of the necessity requirement amounts to violation of the TBT Agreement.
- If the response is yes, then the next question should be whether the necessary measure has been applied in a non-discriminatory manner. In this part of their analysis, the WTO courts should inquire into whether the necessary measure is applied in an even-handed manner across WTO members.

The discussion on non-discrimination, to which we return in what follows, should be sequential to the discussion on necessity. These are not two independent obligations but, rather, one coherent whole. The Appellate Body would simply have to see the TBT Agreement as its current understanding of Article XX of the GATT. There is only one twist: the burden of production of proof stays with the complainant and not the defendant.

One final point regarding the allocation of the burden of proof is probably warranted here. Practice lends support to the argument we have advanced that the Appellate Body is not an umpire shifting the burden of proof every time it is persuaded that a presumption has been raised.[43] Instead, the Appellate Body asks both parties to submit evidence and will judge on the preponderance of the evidence. Still, the first shot should originate in the complainant.

B *Non-Discrimination*

1 Policy Likeness

Market likeness,[44] if at all, is part of the 'esoteric' contemplation that WTO members go through when deciding whether intervention is necessary in the first place. If WTO members approved of the manner in which their own consumers were behaving in the market, and/or if they thought that intervention was not necessary, then they would not have enacted a technical regulation or a standard in the first place. It is policy likeness that matters in the context of the TBT Agreement.

2 Less Favourable Treatment

The Appellate Body should simply ask, assuming two goods are policy-like of course, whether the measure has been applied in an even-handed manner. The exclusivity requirement should be dropped because it is both unwarranted in light of the overall

[43] Horn and Mavroidis, 'Burden of Proof in Environmental Disputes in the WTO: Legal Aspects', 18 *European Energy and Environmental Law Review* (2009) 112.

[44] Neven, 'How Should Protection Be Measured under Art. III of the GATT', 17 *European Journal of Political Economy* (2001) 421; Neven and Trachtman, 'Philippines-Taxes on Distilled Spirits', 12 *WTR* (2013) 297, have expressed similar views on this score.

TBT context as well as almost impossible to prove. There is no need to add a quixotic quest to serve the purpose of the TBT Agreement.

C *International Standards*

Assuming deviation from an international standard that could have been appropriately used, the burden of proof to justify the adopted measures should switch to the party possessing information to this effect – that is, the deviating party.

D Standards

In *US – Tuna II (Mexico)*, the Appellate Body reduced the scope for standards to redundancy. Indeed, if access to a standard is uninhibited, and anyone can profit from, say, a 'dolphin-safe' label, irrespective whether it has complied or not with the statutory requirements, why adopt a standard in the first place. Standards, like technical regulations, are adopted whenever government and private preferences differ. The difference between the two instruments is confined to their regulatory intensity. When regulatory choices are important, technical regulations will be adopted. When issues of lesser interest are at stake, it is time for standards. In our example, allowing for sales of 'dolphin-unsafe' tuna is proof in and of itself that the 'dolphin-safe' label is a standard. It would have been a technical regulation only if 'dolphin-unsafe' tuna could never access the US market.

5 Conclusions

In this article, I have analysed the TBT Agreement as 'completed' through case law. WTO jurisprudence is as disappointing intellectually as it is unfaithful to the intentions of the framers. It is not the law that is wanting. It is case law that has failed. This is good news since all that needs to be done is undo the current case law. It is easy to fix since there is no need to enter a cumbersome renegotiation of the TBT Agreement across 164 heterogeneous partners at a point in time when, as the recent experience in Buenos Aires has amply demonstrated, they find it difficult to agree even on the drafting of a simple ministerial decision with little substantive content.

The Appellate Body, of course, is an agent and not a principal as Article 3.2 of the Dispute Settlement Understanding makes abundantly clear. It must observe the policy space conferred to the WTO by its framers. Its interpretative discretion is what this term suggests; it is an interpretative, as opposed to law-making, function. To be sure, with the exception of its treatment of international standards, the Appellate Body has not affected the policy space committed. It has, on the other hand, turned the test for consistency on its head. Why has this been the case? Largely, because the Appellate Body has consistently failed to understand that the Vienna Convention on the Law of Treaties (VCLT) – its instrument for interpreting not only the TBT Agreement but all of the WTO – is a rule and not rules of interpretation.[45] Eduardo Jimenez de Arréchaga puts it so well when explaining that the reason why Article 31.1 of the VCLT was

[45] Vienna Convention on the Law of Treaties 1969, 1155 UNTS 331.

expressed in one sentence was precisely because the guidance to interpreters was to avoid favouring one of the elements mentioned over the other.[46]

The Appellate Body has managed to do by and large the opposite. It consistently over-emphasizes textualism, its pronouncements to the opposite notwithstanding, and undermines contextual interpretations. To use Joel Trachtman's inimitable expression, the *Oxford English Dictionary* has emerged as its most frequently used source of inspiration.[47] In a similar vein, Henrik Horn and Joseph Weiler, when criticizing the attitude of the Appellate Body towards international standards in its *EC – Sardines* jurisprudence, mention:

> True to their belief in a textualist method of interpretation, out come the dictionaries! The Panel comes armed with Webster. The AB fields its favorite Oxford Shorter. And we let the learned wordsmiths whose dictionary definitions are the most extreme example of understanding language independently of context, and with no reference to object and purpose (i.e. the exact opposite approach to meaning of words which a legal interpreter of international texts should adopt), decide for the WTO the relationship between international standard setting and national administrative procedures.[48]

Dictionaries, however, per construction, privilege interpretations of the widest possible use. This is why they appeal to large categories of users and not to specialists who care about particular uses. Even the most sophisticated dictionaries will refer to only a few contexts, and trade agreements are not high up on that list. GATT-think is not TBT-think, and, as a result, even when the same terms appear in the two agreements, they should be given their proper contextual meaning, as argued above.

Has the recent compliance report solved the problem? The problem has not been solved, and this is a salient feature of this article. The problem will not be solved unless we can legitimately expect predictable outcomes in the future. This will be the case only when a coherent test for consistency has been developed. We quote from Richard Haas: 'One factor … increasing the odds that world order will survive is that it not require talented statesmen, the supply of which is likely to be insufficient. … Individuals of mediocre or poor skills will enter into positions of responsibility.'[49] Respecting proportions, of course, since Haas refers to the world order and here we deal with the case law of the Appellate Body, his point is highly relevant to our discussion. The recent compliance panel did not change anything in terms of the test of consistency with the TBT Agreement and still reached a different position. Most likely, a reasonable panellist would manage to drive this point through. Similar results though can only be guaranteed if the test itself becomes more rational. This is what we tried to propose in this article.

The rationale for our approach is straightforward. The WTO Appellate Body should always ask one question first: Why was an agreement signed? Having provided the response to this question, it will find it much easier to interpret the terms contained therein in line with the objective function of the instrument. It will thus be respecting the mandate of the VCLT as well as its institutional role – that of the agent. It has not done it in the context of the TBT Agreement so far. It is high time for change.

[46] Jimenez de Arréchaga, 'International Law in the Past Third of the Century', 159 *Receuil des Cours* (1978) 1.

[47] Trachtman, 'The WTO Cathedral', 43 *Stanford Journal of International Law* (2007) 127, at 136.

[48] Horn and Weiler, *supra* note 14, at 140.

[49] R.N. Haas, *A World in Disarray* (2016), at 5.