European University Institute

From the SelectedWorks of Lukasz A Gruszczynski

2014

Science and the Settlement of Trade Disputes

Lukasz A Gruszczynski, Institute of Law Studies, Polish Academy of Sciences

Available at: https://works.bepress.com/lukasz_gruszczynski/24/
1. Introduction

Different WTO agreements require, implicitly or explicitly, recourse to science and scientific expertise in order to determine (or assist in the determination of) the legality of national health and environmental measures that impact on international trade. The GATT 1994 stipulates in Article XX that Members may justify a measure that would be otherwise inconsistent with GATT disciplines if such a measure is, among other things, necessary to protect human, animal or plant life or health or related to the conservation of exhaustible natural resources. Science is used in order to establish necessity, i.e. to show the existence of specific risks and to prove the required relation between a measure and an identified risk. Similarly, the TBT Agreement provides that technical regulations shall not be more trade-restrictive than necessary to fulfil various legitimate objectives, including the protection of human health or safety, animal and plant life or health, or the environment. The necessity requirement is again assessed with the help of scientific evidence and technical data. The most elaborate framework is, however, provided by the SPS Agreement, which openly designates science as a criterion for distinguishing between permitted and prohibited measures.

This chapter submits that science has become, in the context of the WTO, an important element (or even a point of reference) in the assessment of domestic health and environmental measures with respect to their compatibility with WTO obligations. Under current practice, WTO panels and the Appellate Body seem to be concerned not only with external legitimacy (i.e. the
absence of alternative less-trade-restrictive measures that could meet the health policy objective), but also with the internal legitimacy of domestic measures (i.e. the existence of a scientific basis). This sharply contrasts with the approach taken under GATT 1947, where panels tended to concentrate only on the first element. This chapter also suggests some reasons that could explain such a shift. In particular, it is argued that, besides the need to guarantee the consistency of different WTO agreements and the increased legalisation and formalization of the WTO dispute settlement process, the legitimizing character of science is also an important factor.

The chapter proceeds as follows. The first part briefly analyses the relevant provisions of the SPS Agreement, a treaty that introduced science and scientific criteria into the domain of international trade law, as well as the corresponding case law. The second part turns to GATT 1994 and the TBT Agreement, and compares the approach of WTO dispute settlement bodies with the practice under GATT 1947. The last part summarizes the previous discussion and elaborates on the reasons behind the new science-based approach of WTO dispute settlement bodies.

2. The SPS Agreement and its scientific obligations

The SPS Agreement establishes certain standards for national SPS regulations. This category is defined by the SPS Agreement as measures adopted for the protection, within the territory of an importing WTO Member, of the life and health of people, animals, and plants from enumerated SPS risks. While the SPS Agreement reiterates traditional GATT principles, such as national treatment or most favoured nation status (with some modifications necessary in the SPS context), it also adds science-based obligations that appear to go beyond the mere elimination of discrimination. On one hand, it recognizes the right of each WTO Member to establish any level of protection deemed appropriate and to take all necessary steps to that end. On the other hand, it requires that all domestic measures which do not conform to international standards (in the majority of cases measures aiming at higher levels of protection than those provided by international standards) to be based on scientific principles and cannot be maintained without sufficient scientific evidence (Articles 3.3 and 2.2). This obligation is further clarified in Article 5.1, which explains that the ‘scientific basis’ needs to take the form of a formal risk assessment. Such an assessment is of scientific character and aims at evaluating the potential (or probability) for adverse effects resulting from the importation of a particular product – e.g. due to the presence
of contaminants and toxins in food or disease-causing organisms in plant and animal products (Annex A(4)). Although the SPS Agreement does not predetermine any minimum magnitude of risk that can trigger a national response, the risk, which is subject to regulation, needs to be ascertainable, i.e. having some indication of potentiality. Consequently, the identification of a theoretical uncertainty relating to the safety of a specific product is insufficient to justify an SPS measure. It is also worth noting that older SPS jurisprudence was not entirely clear (if not inconsistent) about this issue. For example, one of the panels used the concept of ‘negligible risk’, which implies the existence of some threshold limit, and refused to recognize such a risk as a legitimate subject of regulation.

Risk may be expressed both in quantitative and qualitative terms. The quantitative measurement provides information on the probability of the occurrence of an adverse effect (e.g. 1:1,000,000), while the qualitative measurement relates only to the possibility of a causal link (e.g. low or medium), without indicating its likelihood. In any case, risk assessment needs to be specific, meaning that it has to evaluate the particular potential (probability) of harm arising from a specific SPS risk (e.g. evaluation of ‘entry, establishment, or spread of fire blight through [US] apple fruit as a separate and distinct vector’). Consequently, a general discussion of the potentiality/likelihood or assessment of whole categories of hazardous agents is insufficient for the purposes of Article 5.1. The more recent SPS case law, however, has admitted (by making a textual argument that any assessment of risk needs to be performed ‘as appropriate to the circumstances’) that the specificity requirement is not of an absolute character. This means that an assessment may be influenced by methodological difficulties posed by the nature and characteristics of a particular hazard (e.g. a chemical substance). Consequently, in the US/Canada – Continued Suspension dispute, it was not necessary for the EC to show a direct causal link between potential adverse human health effects and residues of oestradiol-17β, because residues of this hormone were only one of the many hormones that humans are exposed to. A more general examination of overall risk resulting from exposure to different hormones, both endogenous and artificial, was considered to be sufficient. This closely corresponds with the finding of the EC – Biotech Products panel, which held that when deciding on a measure Members may account for existing scientific uncertainties to be taken as a consequence of evaluation of risk. According to the Panel, the same risk assessment can equally support very different measures, not only because
of variations in the relevant levels of protection but also due to differing assessments of identified uncertainties.\textsuperscript{11}

The SPS case law is consistent in holding that risk assessment may be based not only on mainstream science, but also on the opinions of scientists taking a divergent view.\textsuperscript{12} In its attempt to eliminate ‘junk’ science as a possible justification for a measure, the Appellate Body held that such opinions ‘must … have the necessary scientific and methodological rigour to be considered reputable science.’\textsuperscript{13} In other words, minority scientific views need to meet a certain epistemic threshold in order to be qualified as legitimate science. In scientific practice this is normally determined through various mechanisms, such as evaluation of a specific claim against more general theories or established scientific views, the logical coherence of a claim, its methodological soundness, or its ability to explain a particular phenomenon.\textsuperscript{14} This means that, at least in theory, the requirement of a scientific basis may be satisfied under the SPS Agreement by one of several rival views.\textsuperscript{15} Having said this, it also needs to be noted that some SPS panels have sought to establish the ‘best science in the field’ rather than verify the scientific quality of arguments put forward by defendants.\textsuperscript{16} Obviously this has the effect of limiting the role of minority scientific opinions. The risk of marginalizing minority opinions is also exacerbated by the specificity requirement discussed above. As noted by Peel, many divergent views are ‘based on the kind of suggestive but not definitive scientific evidence,’\textsuperscript{17} which in turn may result in them having limited value in the eyes of panels.

The SPS Agreement accepts that WTO Members sometimes need to act despite the lack of sufficient scientific evidence (i.e. ‘insufficiency of scientific evidence’ in the language of the agreement)\textsuperscript{18} that would allow them to perform a required risk assessment. The SPS case law has struggled to distinguish this notion from the concept of scientific uncertainty. Although the official position is that the two terms are completely different, with scientific uncertainty not capable of triggering the application of Article 5.7,\textsuperscript{19} a deeper analysis shows that at least some forms of scientific uncertainty have been regarded as instances of insufficiency (e.g. practical immeasurability and indeterminacy).\textsuperscript{20} Overall, the first condition of Article 5.7 was interpreted rather strictly, with some panels refusing to recognize a normative dimension in the assessment of sufficiency/insufficiency (i.e. subjective aspects of a risk assessor’s decision-making that are influenced by the culture of a specific society). In practice this meant that panels considered that
the chosen level of protection could not influence the parameters of evaluation of scientific evidence.\textsuperscript{21} This was an undesirable development. As noted by a group of eminent scholars, ‘evidence deemed reliable enough to generate a sufficient risk assessment in one regulatory context may fail in other contexts because of the different concerns, risk frames, and particular circumstances.’\textsuperscript{22} These concerns were recognized and addressed by the Appellate Body in one of its more recent decisions, where it held that the level of protection sought might play a role in the assessment of insufficiency (e.g. by framing the scope and the methods of risk assessment).\textsuperscript{23} In a similar vein, it recognized that the spatial aspect of insufficiency (i.e., a concurrent existence of insufficiency and a relevant international standard) is independent from the temporal dimension (i.e. scientific developments that take place after the adoption of a standard).\textsuperscript{24}

The remaining part of Article 5.7 requires Members to base their provisional measures on available pertinent data. Moreover, Members are obliged to actively seek additional scientific information in order to perform a more objective assessment of risks, and to review measures within a reasonable period of time.

The SPS Agreement also encourages panels to consult experts on technical and scientific matters, either at the request of a party to a dispute or on its own initiative (Article 11.2). The consultation mechanism is intended to help a panel overcome its epistemic limitations when confronted with complex scientific and technical issues. Although panels may choose, under Article 11, between an advisory experts’ group and individual experts,\textsuperscript{25} in practice all of them have preferred the second option.\textsuperscript{26}

An interesting development that has taken place in the context of the SPS Agreement is the progressive intensification of the applicable standard of review with respect to scientific evidence put forward by defendants (as well as a recent shift in \textit{US/Canada – Continued Suspension} in the opposite direction). In theory, such a standard may vary from \textit{de novo} review to full deference. Under \textit{de novo} review, one body is able to review all the factual determinations (including scientific ones) made by another body and substitute them with its own. At the other end of the spectrum, a fully deferential standard restricts the powers of the reviewing body to monitor procedural compliance (i.e. whether the prescribed procedure(s) were followed), and prohibits any inquiry into the substantive merits of factual determinations. Between these two extremes lie a number of less or more deferential types of review.
The SPS case law adopted the so-called ‘objective assessment of facts’ standard, which in early practice was quite deferential. Subsequent panels, although formally following the ‘objective assessment’ standard, conducted more intrusive reviews of the scientific basis supporting national measures, inquiring into the quality, persuasive force, and correctness of scientific findings made on the domestic level. The reports in *EC – Biotech Products* and *US/Canada – Continued Suspension* were probably the culmination points of this trend, with both panels engaging in an in-depth analysis of the scientific evidence and choosing between competing scientific views. The more recent case law is more ambiguous. The Appellate Body in *US/Canada – Continued Suspension* reverted back to a relatively deferential approach and instructed the panel to concentrate on the coherence of domestic risk assessment and evaluate whether it contained minimum epistemic value. As a consequence, the panel was not expected to assess the correctness of the domestic risk assessment. The subsequent case law (i.e. *Australia – Apples*) further obscured the picture. On one hand, the Appellate Body explained that a panel has to only determine whether a specific claim can be regarded as ‘legitimate’ science (and not the ‘best science’). This implies a rather deferential approach and follows the principles set out in the *US/Canada – Suspension* report. On the other hand however, it accepted a rather intrusive examination into the reasoning included in a risk assessment. Consequently, it is not enough for a panel to inquire whether such reasoning falls within a range that could be considered legitimate by the scientific community. This obviously opens up the possibility of a quite intrusive investigation into the reasoning underlying a risk assessment, including its assumptions, intermediate conclusions, and final determinations.

Overall, the SPS case law has established a relatively strict science-based framework for assessing national SPS measures. Although the system provides some flexibility (especially when it comes to human health risks), a number of standards are formulated in a strict fashion that may unnecessarily constrain the ability of WTO Members to regulate health and environmental risks (e.g. in terms of specificity of risk assessment and conceptualizing insufficiency of scientific evidence as a quasi-objective category). The applicable standard of review still remains relatively intrusive, with a panel entitled to make an inquiry into the scientific basis underlying a national measure (at least with respect to the methodology used in risk assessment).
3. GATT 1994 and the TBT Agreement

The issue of science and scientific evidence is relevant not only under the SPS Agreement, but also in the context of other WTO agreements. Arguably the most important in this context are GATT 1994 and the TBT Agreement. Although other parts of WTO law, such as the TRIPS Agreement or the GATS, may also require an assessment of scientific data, until now there has not been any relevant case law with regard to these agreements. As a consequence, they are not discussed in this chapter.

This section summarizes the major science-related developments that have taken place under both the GATT 1994 and TBT Agreement. However, before moving to the analysis of specific provisions of WTO law, it is worth reviewing the practice under the old GATT 1947.

3.1. The traditional approach under GATT 1947

The relevant GATT 1947 case law is very limited. Genuine scientific questions appeared only once, in Thailand – Cigarettes, a case concerning Thai restrictions on the importation and sale of foreign cigarettes. The US argued that the ban was inconsistent with Article XI of GATT 1947 (i.e. prohibition of quantitative restrictions) and could not be justified by the exception contained in Article XX(b). One of Thailand’s arguments was that the chemicals and other additives contained in US cigarettes (a complainant in the case) might make them more harmful to human health than their Thai counterparts. Moreover, Thailand also submitted that some US cigarettes ‘contained nicotine which was extracted from tobacco leaf, resprayed back into the leaf as part of a process called “reconstituting” the tobacco.’ According to Thailand, this process might have made US cigarettes more addictive, since it could have resulted in the easier inhalation and absorption of nicotine by the bloodstream and the brain.

The Panel held that the measure violated Article XI of GATT 1947. Although it recognized the priority of human health over trade liberalization, it found the measure to be unnecessary because other methods, which were less inconsistent with GATT 1947 (and thus constituting the least restrictive alternative, could have reasonably been used by Thailand to address its health concerns. These included, among others, a ban on the advertisement of cigarettes of both domestic and foreign origin or the creation of a governmental monopoly for the importation and
sale of cigarettes. When deciding the case, the GATT panel took for granted that the Thai measure was backed up by sufficient scientific evidence. At the same time, it did not address the claims of Thailand that US cigarettes were actually more harmful than Thai ones (due to the presence of additives and the specific process used in their production). Any finding in this respect could arguably have influenced the assessment of available alternatives (i.e. whether a controlled use or a ban on advertisement was really a feasible alternative). On one hand, the Panel’s approach appeared to be very deferential by assuming the existence of sufficient scientific evidence supporting the measure. On the other hand however, it was also very intrusive as it exercised its discretion in substituting the judgment of a national government concerning available alternatives with its own views, without examining the internal legitimacy of alternative measures.

Goh identifies a number of reasons which could explain the Panel’s approach. In particular, he argues that the trade-effects test used by the Panel (which concentrated on the external legitimacy of the measure) corresponded closely with the character of GATT 1947. As a trade instrument ‘it encouraged the development of practical trade-effects tests in examining the legitimacy of measures, as opposed to internal or “scientific” tests.’ 37 It seems that GATT panels in general were rather reluctant to develop any new tests especially when faced with politically controversial disputes. As Wirth notes, GATT panels tended to be deferential to any substantive determinations (including those which involved scientific elements) made by national governments, and concentrated more on procedural compliance. 38 For example, in the Tuna cases 39 - another health/environment-related disputes - the panels avoided any discussion of the scientific merits of the US importation ban (e.g. whether it was necessary and whether possible alternatives might be sufficiently effective), 40 and concentrated instead on the problem of the extra-territorial application of the US measure and its unilateral character. 41

4.2. GATT 1994

One would expect to see the same approach under GATT 1994. Despite the fundamental institutional changes that took place with the creation of the WTO and the adoption of various new agreements, the text of GATT 1994 remained the same as that of GATT 1947. In addition, the newly-created Appellate Body from the outset recognized GATT 1947 jurisprudence as relevant in the interpretation of WTO agreements. Indeed the first case US – Gasoline appeared to confirm
the previous approach. The Panel simply assumed the existence of sufficient scientific evidence underlying a measure, noting that air pollution, and in particular ground-level ozone and toxic substances, presented health risks to humans, animals, and plants. Thus the panel concentrated its analysis on available alternatives (non- or less restrictive), but it did not discuss their scientific plausibility. Although the Appellate Body introduced a number of important changes to the panel report, it did not comment on the Panel’s treatment of science and scientific evidence.

The next case with an environmental dimension (US–Shrimp) was more ambiguous. Although none of the parties requested consultations with experts, the panel nevertheless decided to consult them on a number of scientific issues, explaining its decision in this respect by the amount of scientific documentation submitted by both parties. In particular, it extensively inquired into two research areas: approaches to sea turtle conservation, and the habitat and migratory patterns of sea turtles. On its face, this significantly differed from the approach of the GATT US–Tuna panel, despite many factual similarities between the two cases. As a consequence, one might expect that science played an important role in deciding the dispute. However, this was not the case. In the end the Panel did not engage in any discussion on the underlying science (e.g. as to the effectiveness of specific turtle conservation policies), and avoided all scientific questions by starting its analysis from the chapeau of Article XX. After finding that the US measure amounted to arbitrary and unjustifiable discrimination (the traditional GATT test), it stopped short of examining specific subparagraphs of the general exception, where scientific issues could be of prime importance. The Appellate Body changed the sequence of analysis, but in its examination under subparagraph (g) it also did not go into the details of the internal (i.e. scientific) legitimacy of the measure. Although the Appellate Body found that sea turtles could qualify as exhaustible natural resources and that the US measure was related to the conservation of such resources, it came to these conclusions not on the basis of expert opinions or scientific evidence, but by referring to various international environmental instruments (e.g. the Convention on International Trade in Endangered Species of Wild Fauna and Flora). The rest of its analysis was not concerned with the scientific aspects of the US measure, but instead procedural and transparency questions played a central role.

The approach of the WTO dispute settlement bodies changed with the famous EC–Asbestos case, which concerned the French ban on the importation of certain asbestos products due to risks
to human health posed by exposure to asbestos fibres. Both the Panel and the Appellate Body engaged in a detailed examination of the internal legitimacy of the measure, including the nature and degree of the risk being addressed. At the same time, there were some important differences between the reasoning of the Panel and the Appellate Body. The approach of the former was more market-oriented, as its analysis of the likeness of different products (in casu chrysotile fibres (asbestos) and cellulose fibres/PVA/glass fibres) under Article III:4 of GATT 1994\(^5\) concentrated on their competitive relationship. As a consequence, the Panel decided on the issue of likeness in terms of narrowly understood competitive substitutability and did not really consider as relevant the capacity of asbestos to cause cancer.\(^5\) Only after concluding that the French ban was inconsistent with the requirement of Article III:4 of GATT 1994 did it find the scientific evidence pertinent for its analysis under Article XX. This included two specific issues: (i) the existence of a risk to human health for the purpose of determining necessity and (ii) the existence of other measures which are either consistent or less inconsistent with GATT 1994 and achieve the same level of protection.\(^5\) The approach of the Appellate Body, while also concentrating on the internal legitimacy of the measure, was entirely different. The Appellate Body, similar to the Panel, saw the assessment of likeness under Article III:4 in terms of competitive relation, but it also considered the potential health risks posed by one of the products (which was shown by scientific evidence) to be a relevant factor in the final determination as whether two products are like or not.\(^5\) In particular it found that carcinogenicity constituted one of the defining elements of the physical properties of the products involved, and that it might have influenced consumer preferences.\(^5\)

The approach adopted in *EC–Asbestos* bears many similarities to the SPS standards. Although the panel expressed its reservation about following SPS case law,\(^5\) in fact both the Panel and the Appellate Body relied on various mechanisms developed in the context of the SPS Agreement. Both were also concerned with the internal and external legitimacy of the measure, thus going beyond the traditional GATT 1947 approach. In addition such a change forced the Panel to consult with experts on a number of scientific issues (e.g. the existence of risk resulting from exposure to chrysotile fibres and the efficiency of alternatives identified by Canada in achieving the French health objectives). The Appellate Body recognized, in line with the SPS Agreement standards, that risk as such did not need to be assessed in quantitative terms and that qualitative measurement was also sufficient under GATT 1994.\(^5\) In addition the measure could be supported by not only
mainstream scientific research, but also using minority views. As far as the latter category is concerned, a certain quality threshold was introduced, as minority scientific opinions had to come from qualified and respected sources.\textsuperscript{57} This corresponded closely with the treatment of minority views under the SPS Agreement. The applicable standard of review was quite deferential and followed the initial approach of the WTO dispute settlement bodies under the SPS Agreement. In this context, the Panel explained that ‘it is not its function to settle a scientific debate, not being composed of experts in the field of the possible human health risks posed by asbestos. Consequently, the Panel does not intend to set itself up as an arbiter of the opinions expressed by the scientific community.’\textsuperscript{58} Its limited role only required assessing whether the evidence was sufficient to conclude that there was a human health risk, and whether the measure was necessary to achieve France’s legitimate goal (and not to determine whether the risk exists ‘objectively’, e.g. according to the best available science).

The subsequent health/environment-related case law has followed the standards established by \textit{EC – Asbestos}. The panel in \textit{Brazil – Retreaded Tyres}\textsuperscript{59} found the Brazilian measure to be a quantitative restriction in violation of Article XI of GATT 1994, and examined the measure under Article XX. The Panel clarified that a product which is regulated does not need to pose a direct risk in order to fall within the ambit of Article XX(b). This was an important development since in \textit{EC – Asbestos} chrysotile fibres posed direct and immediate risks to human life and health. The situation in \textit{Brazil – Retreaded Tyres} was different, as retreaded tyres as such did not give rise to any risk to human, animal or plant life and health. It was only the accumulation of inappropriately disposed waste tyres which posed risk. Thus, the Panel was satisfied with the fact that there was a relationship between the risks arising from the accumulation of waste tyres and the ban on importation of retreaded tyres. As far as the substance of the measure was concerned, the Panel concentrated on two issues. First, it examined whether there was a risk to human, animal or plant life or health and, if so, whether the objective of the import ban was to reduce such risk. Although the Panel did not consult experts, in its analysis it frequently referred to scientific evidence. Interestingly, the Panel was satisfied with the evidence, which merely suggested a correlation between the spread of different diseases and the accumulation of used tyres. It did not require any specific evaluation of the risk posed by the improper management of used tyres in Brazil. It accepted evidence of general nature (e.g. the Basel Convention Technical Guidelines on the
Identification and Management of Used Tyres), including indirect evidence relating to practices in other countries (e.g. the European Union, UK).\textsuperscript{60} The Panel also explicitly refused to recognize the specificity requirement (i.e. evidence of the actual negative health effects of tyre fires within Brazil).\textsuperscript{61} This was definitively a much more liberal approach than the one required by the SPS Agreement. Second, the Panel analyzed the necessity of the measure, relying on the process of weighing and balancing different factors. Scientific evidence appeared to be of prime importance in this context, both for the examination of contribution of the measure to realization of the end pursued (i.e. reduction of the risks to human, animal and plant life and health) and the availability of alternatives measures (i.e. whether an alternative is a feasible option). Although the Appellate Body reversed various findings of the Panel, it upheld the entire necessity analysis, including the examination of possible alternatives to the Brazilian measure, its contribution to achievement of the regulatory objective, as well as the weighing and balancing of various relevant factors.\textsuperscript{62} As noted above, the necessity test was that part of the Panel’s analysis which was dominated by scientific analysis.

3.3 The TBT Agreement

The TBT Agreement, similar to its SPS counterpart, openly refers to science. To start with, Article 2.2 requires WTO Members to ensure that their ‘technical regulations are not prepared, adopted or applied with a view to or with the effect of creating unnecessary obstacles to international trade.’ For this purpose, they cannot be ‘more trade-restrictive than necessary to fulfil a legitimate objective, taking account of the risks non-fulfilment would create.’ The list of legitimate regulatory objectives is open-ended and includes protection of human health or safety, animal or plant life or health, as well as the environment. When assessing the risks of non-fulfilment, WTO Members are expected to take into account, among other things, available scientific evidence. In addition, Article 14.2 explicitly confirms the right of a panel, in the context of the TBT Agreement, to consult experts on questions of a technical nature.

Considering this explicit reference to science, one should not be surprised to find analyses of the internal legitimacy of TBT measures under Article 2.2. Two recent cases may serve as good examples. The \textit{US – Clove Cigarettes} dispute concerned the US ban on production and sale of flavoured cigarettes (including clove cigarettes), with the exception of the menthol version. The
aim of this prohibition was to reduce youth smoking by eliminating cigarettes that, by masking the regular taste of tobacco (which may be unpleasant for beginners), are particularly appealing to young people. Indonesia claimed, among other things, that the ban was discriminatory (Article 2.1) because it treated menthol (mainly domestically produced in the US) and clove cigarettes (all of which were imported) differently, and furthermore that the ban was unnecessary (Article 2.2). Since the case did not concern the relative toxicity of clove, menthol, and regular cigarettes (due to the presence of some specific additives), the Panel was not expected to look at the scientific evidence that would support such a contention. However, under the necessity analysis, the Panel was confronted with the question of the effectiveness of the US measure (i.e. whether banning clove cigarettes would contribute to the achievement of the regulatory objectives sought by the US). In answering this question, the Panel engaged in a detailed analysis of the scientific evidence and ultimately found that extensive scientific research supported the conclusion that banning clove and other flavoured cigarettes could contribute to a reduction in youth smoking. In the context of its analysis, the Panel also indirectly recognized that minority scientific opinion could constitute a sufficient basis for TBT measures. On the other hand, the issue of internal legitimacy was not so visible in the Panel’s discussion of available alternatives that might have been employed by the US. This was, however, more connected with the structure of Indonesia’s argument rather than an axiomatic stance taken by the Panel. Since Indonesia did not appeal the Panel’s findings under Article 2.2, the Appellate Body did not have the opportunity to review this part of the report.

A similar approach was taken in a subsequent TBT dispute, US – Tuna II (Mexico). The case concerned the US measure establishing the conditions for use of ‘dolphin safe’ labels on tuna products. Mexico argued that the measure was incompatible with certain TBT provisions, in particular Article 2.1 (because Mexican tuna products had received less favourable treatment as compared to products originating in other countries, including the US), as well as Article 2.2 (as being more trade-restrictive than necessary to fulfil the legitimate objective(s) sought by the US). The Panel, when assessing the measure under Article 2.2, enquired whether the US measure fulfilled the stated objective (i.e. protection of consumers by ensuring that they are not misled as to the quality of tuna products and protection of dolphins by discouraging certain fishing techniques that harm them), and whether there were other less trade-restrictive alternatives. In this context, it analysed a large amount of scientific data on the impact of purse seine operations
on the health of wild aquatic mammals, dolphin mortality rates associated with fishing techniques other than what is referred to as ‘setting on’ dolphins, and risks of dolphin mortality or injury arising from tuna fishing operations outside the Eastern Tropical Pacific Region. The Panel’s examination of available alternative measures was also partially based on its previous scientific determinations. All those endeavours were aimed at assessing the internal (i.e. scientific) legitimacy of the US measure. The Panel also touched upon the standard of review applicable to scientific evidence. Similar to the approach taken under GATT 1994, it was interested in determining whether specific claims were sufficiently substantiated by the relevant party rather than whether they represented the best available science.

The Appellate Body reversed the Panel’s findings under Article 2.2. It explained that the relevant test required the Panel to determine: (i) the existence of a legitimate objective(s); (ii) the degree to which a contested measure contributes to such an objective(s); (iii) the trade-restrictiveness of a measure and (iv) the nature of the risks and the gravity of the consequences arising from non-fulfilment of the objective(s) pursued by the Member. The latter three elements are used to establish whether a measure is more trade-restrictive than necessary. The Appellate Body also endorsed a comparison of a contested measure with (reasonably) available alternatives as a conceptual tool for ascertaining its relative necessity. Although it disagreed with the Panel that the alternative measure identified by Mexico (i.e. co-existence of the US and AIDCP label) would contribute to the achievement of the regulatory objective to the same extent as the existing US measure, this finding did not alter the approach with respect to its treatment of scientific data. In the context of the analytical framework proposed by the Appellate Body, scientific evidence still seems to be highly relevant to a determination of whether there is a genuine risk that needs to be addressed, the nature of such risk, the contribution of a measure to the achievement of the objective, and the feasibility of identified alternatives.

4. Scientific standards and health/environment-related trade disputes

As discussed above, science has become a central benchmark for the assessment of SPS measures. Domestic regulations are tested against science-based criteria with respect to their internal legitimacy. Having said this, it is not clear what drafters had in mind when they decided to introduce such a novel mechanism into the trade regime. A part of the literature argues that
science, under the SPS Agreement, is used as a sophisticated instrument for the detection of protectionist regulations. As explained by Sykes, the requirement of scientific justification ‘aids in motive review, and helps to sort regulations between those that are protectionists and those that seek to promote some legitimate, non-protectionist regulatory objective.’ The scientific obligations, however, seem to go further than a mere detection and elimination of protectionism. Science acts more as a mechanism aimed at the elimination of all ‘unnecessary’ (assessed against some objective standard) restrictions on international trade. Consequently, scientific criteria under the SPS Agreement aim at eliminating obstacles to international trade, irrespective of whether they are motivated by protectionism. Irrespective from the motives, the SPS Agreement is predominantly concerned with the internal legitimacy of national measures, while assessment of external legitimacy is of secondary importance (e.g. under Article 5.6). The TBT Agreement is similar. The scope of the TBT Agreement, and particularly Article 2.2, seem to go beyond the mere eradication of protectionism and require that domestic technical regulations maintain a certain standard of internal rationality. Although the GATT 1994 remains preoccupied with the problem of discrimination, nevertheless science plays a more important role than under the old rules.

Analysis of the relevant GATT 1994 and TBT case law further reveals a number of similarities to the SPS Agreement. As noted above, science is an important element in all health and environment-related trade disputes, irrespective of which agreement they are brought under, with WTO dispute settlement bodies inquiring into the internal legitimacy of the measures they examine. In line with the SPS Agreement, the existence of risk can be shown both qualitatively and quantitatively; domestic measures may be based on either majority or minority scientific opinions (if they come from qualified and respected sources); there is no requirement of minimum magnitude of risk (although the risk needs to be ascertainable); and the standard of review applied to scientific determinations at the domestic level remains relatively deferential (which corresponds with at least a part of the SPS jurisprudence). Of course there are still important differences between agreements. For example, there is no obligation, neither under the TBT Agreement nor GATT 1994, to have a risk assessment that would support a national measure. Similarly, the specificity requirement is less stringent. This eventually makes the applicable science-based tests under both agreements less demanding. Nevertheless, considering the earlier GATT approach (and
despite the differences in the language employed in each of the treaties), the change in overall approach is significant. One may therefore wonder about the reasons behind this shift in the practice of the WTO dispute settlement bodies.

In part this development is probably connected with the overall scientification of various spheres of human activity. This is also reflected in the law, both domestic and international, as well as in the settlement of disputes. Consequently, the shift that occurred in the context of WTO law (in the sense of creating new SPS/TBT rules but also modifying the character of the inquiry under GATT rules) should be seen as a part of the broader process that has taken place in a variety of areas. Another possible reason is the progressive legalization and formalization of the trade dispute settlement system. The WTO, as compared to the old diplomatic-like GATT system, is more legalistic and has a formalized dispute settlement mechanism. This requires the parties to a dispute to elaborate on their legal arguments and comprehensively substantiate all factual claims. In the case of health and environmental disputes, a reference to science seems to be indispensable. Panels also tend to produce longer reports that attempt to fully address all the arguments (including all factual aspects) put forward by the parties, as well as to explain the reasoning behind their interpretative choices and factual conclusions. This is partially because of fear of being overruled by the Appellate Body, and partially because panels act as sole fact-finders in the WTO dispute settlement process (without the possibility to remand a case if the Appellate Body finds a factual determination to be insufficient). The need to ensure the overall consistency of WTO case law is another important factor. Since the SPS Agreement provides the most elaborate set of rules for trade measures that have a health or environmental dimension, it is only natural to treat its standards as a kind of ‘guideline’ as to how to approach similar situations (i.e. measures addressing health or environmental risks) under other agreements. It should also be noted that the consistency of case law impacts on the overall legitimacy of the dispute settlement system as a whole, as it increases the predictability of decisions and certainty in the law, thus reducing arbitrariness.

Another factor that may explain the shift in GATT 1994 case law (as well as the introduction of scientific criteria to the SPS and TBT Agreements) relates to the legitimizing nature of science. GATT 1947 was predominantly concerned with tariff barriers and quotas. Its success in reducing these types of obstacles has shifted the attention of countries to market access problems caused by internal measures (i.e. non-tariff barriers). In consequence, potential disputes under WTO law have
become not only more complex (than tariff barriers and quotas), but also more politically contentious.\textsuperscript{85} This has put the WTO dispute settlement bodies under considerable pressure from WTO Members. Both the Appellate Body and panels are accountable to the states which created the WTO system and delegated some of their powers. However, due to the principle of negative consensus WTO Members have only a limited control over the outcomes of disputes. This forces both panels and the Appellate Body to ensure that their decisions are recognized as legitimate and as within the scope of their delegated powers.\textsuperscript{86}

The recourse to science and scientific expertise may thus be seen as a way of improving the legitimacy of the decisions rendered by dispute settlement bodies. As noted by Peel, ‘[i]n emerging areas of international law concerned with the regulation of risk, expertise based on scientific and technical knowledge is typically viewed as a plausible basis for legitimating the growing authority exercised by relevant international rules.’\textsuperscript{87} Science can, therefore, be regarded as a kind of an objective and value-free benchmark against which national measures can be tested and their legality assessed. Deciding disputes on a technical level, with science playing the central role in the assessment of the legality of national measures, depoliticizes (at least on its face) a particular trade controversy. Such an approach reduces the relevance of political/policy considerations in the dispute settlement process – a specific issue is decided against the objective criterion of science and does not require making difficult normative decisions - while reference to the ‘higher’ rationality of science shields dispute settlement bodies against the criticism from the governments of WTO Members. In other words, reliance on science can be seen as a form of denial of any policy-making role on the part of dispute settlement bodies and a defence mechanism against charges that they, acting as agents, have exceeded their delegated powers.\textsuperscript{88} As rightly summarized by Shaffer ‘for legal institutions, it is sometimes thought that the expertise of science can help to resolve competing claims in a neutral manner, evade normative judgments with their distributional implications, and legitimize legal decisions through the authority of science.’\textsuperscript{89} It should also be noted that the WTO, as a relatively young international organization, is particularly vulnerable to the accusation of overstepping its mandate. Its limited institutional record, combined with uncertainties as to the exact scope of rights and obligations arising from the WTO agreements,\textsuperscript{90} produces a situation whereby both panels and the Appellate Body feel a need to be very careful in establishing and maintaining their authority over trade disputes. This is particularly important
when they are confronted with highly sensitive issues involving human health or environmental protection, matters which are at the core of the sovereign prerogatives of each national state.

One may rightly ask whether science has fulfilled the above expectations. Two relatively recent rulings (EC – Biotech Products and US/Canada – Continued Suspension) show that this is not necessarily the case. Disputes that are highly politicized remain contentious irrespective of the criteria that are used to resolve them.\(^9\) Parties tend to disagree about what science actually says in a particular case, highlight or downplay existing uncertainties, and use the complexities involved to further their agendas. Although this is particularly visible in the context of the SPS Agreement where science occupies a special place, real disagreements about specific scientific claims can also occur under GATT 1994 and the TBT Agreement (compare EC – Asbestos). At the same time, one has to admit that in less controversial cases (e.g. Australia – Salmon, Japan – Agricultural Products II, US – Clove Cigarettes) science has indeed served as a very useful tool in assisting panels in their assessment of domestic measures.

The inability of science to resolve some trade disputes is probably connected with its nature. Science, as used in assessment of risk, is not entirely objective criterion. The construction of scientific knowledge is a social process that is culturally dependent,\(^9\) and the assessment of risks relies on a number of normative non-scientific approximations and assumptions (e.g. due to gaps in the knowledge, with some of them being irreducible). This means that science is not capable of being a fully objective benchmark in international trade disputes. In consequence, the reliance on science and scientific expertise, although sometimes helpful, may also sometimes fail to improve the legitimacy of some WTO dispute settlement decisions.

---

\(^1\) For the list of relevant risks, see Annex A(1) of the SPS Agreement.

\(^2\) Relevant international standards are those which are promulgated by the Codex Alimentarius Commission (food safety), World Organization for Animal Health (animal safety) or under the framework of the International Plant Protection Convention (phytosanitary safety). It is also possible for the SPS Committee to identify other organizations as relevant standard-setting bodies.

\(^3\) Appellate Body Report, EC – Hormones, para. 186.


13 Appellate Body, *US – Continued Suspension*, para. 591.


18 On the relationship between disciplines of Articles 2.2/5.1 and 5.7, see Gruszczynski, op. cit., pp. 178-185.


20 *Cf. also*, Gruszczynski, op. cit., pp. 188-191.


23 Appellate Body, *US – Continued Suspension*, paras. 685-86.


For example, scientific evidence could be relevant in the context of Article 20 of the TRIPS Agreement. In particular, science may be helpful in deciding whether a specific encumbrance of a trademark use is justifiable. This is actually one of the issues in the currently pending case relating to the Australian plain packaging law (cf. Australia – Tobacco Products and Packaging (Ukraine), WT/DS434).

The relevant part of Article XX(b) provides that ‘[s]ubject to the requirement that such measures are not applied in a manner which would constitute a means of arbitrary or unjustifiable discrimination between countries where the same conditions prevail, or a disguised restriction on international trade, nothing in this Agreement shall be construed to prevent the adoption or enforcement by any contracting party of measures: (...) (b) necessary to protect human, animal or plant life or health.’

Another argument put forward by Thailand was that foreign cigarettes were particularly attractive to local consumers and could contribute to the increased consumption of tobacco products. In this context, Thailand also argued that various marketing techniques used by foreign companies would have the same effect.


Note that Mexico claimed in the first Tuna dispute that dolphins could be regarded as exhaustible natural resources only if this were shown by scientific evidence (para. 3.44), while the US argued that ‘if a party’s measures were based on scientific information evaluated using recognized scientific approaches, a dispute settlement panel in the GATT should not substitute its own judgment for that of the contracting party whose measure is challenged.’ The Panel did not address either of these arguments.

GATT Panel Report, US – Tuna (Mexico), para. 5.28 (finding that the US did not demonstrate that ‘it had exhausted all options reasonably available to it to pursue its dolphin protection objectives through measures consistent with the General Agreement, in particular through the negotiation of international cooperative arrangements’).

Panel Report, US – Gasoline, para. 6.21 (note that the approach of the panel might have been influenced by the arguments used by Brazil and Venezuela).


Ibid, para. 5.1.

Ibid, para. 7.29.

Subparagraph (g) concerns measures ‘relating to the conservation of exhaustible natural resources if such measures are made effective in conjunction with restrictions on domestic production or consumption.’

48 Ibid, para. 142.

49 J. Peel, *Science and Risk Regulation in International Law*, Cambridge: Cambridge University Press, 2010, p. 272. Note that to some extent the approach of the Appellate Body could have been influenced by the wording of the relevant provisions. Both cases (*US – Gasoline* and *US – Shrimp*) were decided under subparagraph (g), which only requires that a measure is related to a regulatory objective. This is a weaker requirement than the necessity test envisaged by subparagraph (b), and arguably it gives more flexibility to dispute settlement bodies as to choice of the appropriate methodology.

50 The relevant part of Article III:4 of the GATT 1994 provides that: ‘The products of the territory of any contracting party imported into the territory of any other contracting party shall be accorded treatment no less favourable than that accorded to like products of national origin in respect of all laws, regulations and requirements affecting their internal sale, offering for sale, purchase, transportation, distribution or use.’


52 Ibid, para. 8.179.


54 Ibid, paras. 113-114.


57 Ibid, para. 178.


60 Ibid, paras. 7.61, 7.68 and fn. 1118.

61 Ibid, para. 7.77.


63 Panel Report, *US – Clove Cigarettes*, paras. 2.7-2.8.

64 Ibid, para. 3.1.

65 Ibid, para. 7.154.

66 The Panel also discussed scientific evidence in the context of Article 2.1 in order to establish whether clove and menthol cigarettes were like products (e.g. para. 7.229). Although the Appellate Body reversed the panel with respect to some of its findings under this article, this did not concern the treatment of scientific data.

67 Ibid, para. 7.401 and fn. 715 (finding that there was scientific consensus on the connection between flavoured cigarettes and smoking rates in younger groups of consumers, but also referring to the relevant SPS law on minority scientific opinions).


69 Panel Report, *US – Tuna II (Mexico)*, para. 3.1.

70 Ibid, para. 7.473.

71 Ibid, para. 7.496.
Ibid, para. 7.517 and subsequent.

Ibid, para. 7.529.

Ibid, para. 7.504.


A label that was created within the context of the Agreement on International Dolphin Conservation Program (1998).


Goh, op. cit., p. 447.

Article 5.6 provides: ‘(...) when establishing or maintaining sanitary or phytosanitary measures to achieve the appropriate level of sanitary or phytosanitary protection, Members shall ensure that such measures are not more trade-restrictive than required to achieve their appropriate level of sanitary or phytosanitary protection, taking into account technical and economic feasibility.’


But, see the Appellate Body Report, US – Clove Cigarettes, para. 96 (noting that ‘the balance set out in the preamble of the TBT Agreement between, on the one hand, the desire to avoid creating unnecessary obstacles to international trade and, on the other hand, the recognition of Members’ right to regulate, is not, in principle, different from the balance set out in the GATT 1994, where obligations such as national treatment in Article III are qualified by the general exceptions provision of Article XX’).


Since the WTO is an intergovernmental organization, the Appellate Body and panels arguably tend to legitimize their judgments vis-à-vis governments rather than other actors.


Picciotto, op. cit., p. 481 (discussing this argument in the context of legalistic formalism used by the Appellate Body).

WTO agreements can be seen as a type of unfinished contract. Specific provisions are formulated in very general and sometimes ambiguous language, while the system as such has numerous gaps that have to be filled in by the WTO dispute settlement bodies when interpreting and applying WTO law.
