

European University Institute

From the Selected Works of Lukasz A Gruszczynski

2013

Standard of Review of Health and Environmental Regulations by WTO Panels

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



Available at: https://works.bepress.com/lukasz_gruszczynski/12/

[*Research Handbook on Environment, Health and the WTO*, Geert Van Calster, Denise Prévost (eds.), Edward Elgar Publishing, Cheltenham: 2013 (forthcoming)]

Standard of review of health and environmental regulations by WTO panels*

Lukasz Gruszczynski

1. Introduction

The issue of the applicable standard of review in health and environment-related trade disputes has recently become prominent in scholarly discussions. While earlier research tended to concentrate on specific substantive requirements of WTO law,¹ more recent scholarship has turned its attention to this specific procedural question. This shift is most probably a result of developments in WTO case law. The applicable standard of review had a direct impact on the outcome of the *EC – Biotech Products*² dispute. It was also one of the major issues in the Appellate Body ruling in *US/Canada - Continued Suspension*³ and in the more recent *Australia – Apples* case.⁴

* Some parts of this chapter draw on my earlier article: (2011), *How Deeply We Should Go? In a Search of Appropriate Standard of Review in the SPS Cases*, *European Journal of Risk Regulation*, 2(1), 111-114.

¹ There are important exceptions, cf. e.g., Christoforou, 1999-2000; Button, 2004.

² Panel Report, *European Communities – Measures Affecting the Approval and Marketing of Biotech Products*, WT/DS291/R, WT/DS292/R, WT/DS293/R, Add.1 to Add.9, and Corr.1, adopted 21 November 2006, DSR 2006:III-VIII, 847.

³ Appellate Body Report, *United States – Continued Suspension of Obligations in the EC – Hormones Dispute*, WT/DS320/AB/R, adopted 14 November 2008, DSR 2008:X, 3507 and Appellate Body Report, *Canada – Continued Suspension of Obligations in the EC – Hormones Dispute*, WT/DS321/AB/R, adopted 14 November 2008.

The latter report demonstrated that the standard of review, at least in the context of the WTO Agreement on the Application of Sanitary and Phytosanitary Measures (SPS Agreement),⁵ remains ambiguous and will most probably generate further controversies in future WTO disputes.

This chapter analyses the existing WTO case law in order to determine the basic parameters that characterize the applicable standard of review in health and environment-related trade disputes. This should allow a critical assessment of the current practice, identifying both its advantages and disadvantages. The analysis presented here is, however, restricted in two ways. First, the chapter is primarily interested in the standard of review applicable to domestic factual determinations rather than to the legal interpretation advanced by the WTO Members or degree of scrutiny exercised by the Appellate Body over panels' findings. Second, the analysis is limited to the SPS Agreement, and specifically to those provisions which require the review of scientific evidence. While other WTO agreements, such the General Agreement on Tariffs and Trade 1994 (GATT 1994) or the Agreement on Technical Barriers to Trade (TBT Agreement), may be relevant when assessing national health and environmental measures, the problem of applicable standard of review under those agreements remains either a secondary or an abstract issue. The relevant cases law under the GATT 1994 (e.g. *EC – Measures Affecting Asbestos and Asbestos-Containing Products* or *Brazil - Measures Affecting Imports of Retreaded Tyres*) did not require panels to make any complex factual determinations. The existence of risk

⁴ Appellate Body Report, *Australia – Measures Affecting the Importation of Apples from New Zealand*, WT/DS367/AB/R. adopted 17 December 2010.

⁵ Agreement on the Application of Sanitary and Phytosanitary Measures, 1867 UNTS 493, signed on 15 April 1994.

related to the use of asbestos products was rather uncontroversial and supported by significant scientific evidence. The same is true for risks of mosquito-borne diseases (e.g. dengue or yellow fever) connected with accumulation of waste tyres.⁶ On the other hand, any discussion on the standard of review applied to scientific evidence under the TBT Agreement remains theoretical, as until now no case has been decided which would be relevant to the issues addressed here.

This chapter is organized as follows: The first part introduces the concept of standard of review and discusses its different meanings. The second part focuses on the SPS Agreement, briefly describing its basic disciplines and analysing in some detail the applicable standard of review in the early SPS case law. The third part addresses the two most recent rulings (i.e. *US/Canada – Continued Suspension* and *Australia – Apples*), both of which extensively elaborated on the applicable standard of review. The chapter concludes that the WTO dispute settlement bodies have failed to articulate a clear and operable model, and their jurisprudence remains ambiguous or sometimes even contradictory.

2. The standard of review in WTO dispute settlement practice

Standard of review is conventionally understood as the level of scrutiny that is applied by a superior body (a court or a higher administrative authority) over a decision taken by a lower body that is subject to review. Depending on the one's perspective, the standard of review can be therefore defined as 'the degree of deference or discretion that the court accords to legislator or regulator' or 'degree of intrusiveness or invasiveness into the legislator's or regulator's decision-making

⁶ Note also that none of the parties in those disputes argued that any specific standard of review should have been applied by the WTO panel.

process.⁷ Consequently, standard of review determines the extent of discretionary powers enjoyed by a lower body (i.e. lower court or administrative authority) in making certain determinations. In theory, the standard of review may range from *de novo* review to full deference. Under *de novo* review, a superior body is able to review all the determinations made by an inferior body and substitute them with its own. A fully deferential standard restricts the reviewing powers of a superior body to procedural compliance (i.e. whether prescribed procedure was followed) and bars review of the substance. Between these two extremes there are a number of less or more deferential/*de novo* types of review. In practice they appear under different names, such as the ‘arbitrary and capricious’ standard, ‘clearly erroneous’ standard, or ‘reasonable deference’.⁸

The concept of standard of review is common to many national jurisdictions, including all major continental and Anglo-Saxon systems. In a national legal context, it serves as a mechanism for allocating the power between different branches of government (i.e. executive and judicial). A deferential standard favours the body that takes an initial decision (e.g. executive) while *de novo* review introduces additional checks by another body (e.g. judicial). Just as there is no optimal and one-size-fits-all model for the distribution of powers within a state, so too there is also no ideal and universal standard of review. The level of intrusiveness varies from jurisdiction to jurisdiction, reflecting local particularities or current preferences of society - e.g. a need to guarantee a greater oversight by courts over activities of administrative agencies.⁹

⁷ Bohanes and Lockhart 2009, p. 379.

⁸ For an overview of the approaches in US courts see: Strauss, Rakoff & Farina, 2003, p. 902.

⁹ Bohanes and Lockhart 2009, p. 380.

On the international level, the standard of review fulfils a similar function. It determines, alongside the substantive obligations, the distribution of powers between national governments and international bodies. *De novo* review transfers a power to international level at the expense of prerogatives of domestic governments. A deferential standard of review has the opposite effect, empowering national bodies and limiting the competences of international authorities. As noted by one scholar in the context of WTO rules: ‘granting greater deference to the decisions of the state is equivalent to increasing the substantive power of the state to impact trade.’¹⁰ For example, if a WTO panel has only limited competence to re-evaluate scientific evidence which constitutes a basis for a national SPS measure (e.g. whether growth promotion hormones in cattle increase the risk of cancer for humans consuming beef), national authorities gain a wider regulatory freedom. They may evaluate and assess scientific data in a rather unconstrained way, reflecting local preferences and particularities. On the other hand, if the scientific support for a municipal trade measure is reviewed afresh, such measure may be regarded as unjustified if a panel comes to different conclusions than a national government. This aspect (i.e. distribution of powers) was clearly recognized by the Appellate Body in *EC - Hormones*, when it stated: ‘the standard of review (...) must reflect the balance established [in WTO law] between the jurisdictional competences conceded by the Members to the WTO and the jurisdictional competences retained by the Members for themselves.’¹¹

Before going further, one should also conceptually distinguish between the

¹⁰ Guzman 2008, p. 4.

¹¹ Appellate Body Report, *EC Measures Concerning Meat and Meat Products (Hormones)*, WT/DS26/AB/R, WT/DS48/AB/R, adopted 13 February 1998, DSR 1998:I, 135, para. 115.

standard of review applied to factual determinations, and that applied to legal determinations. The first category relates to the review of factual findings made by a body whose decision is subject to review (i.e. factual determinations underlying regulatory decision of a WTO Member), and will be discussed in more detail below. The second category is concerned with the legal interpretations advanced by such a body. It is enough to mention that in the context of the WTO, both the Appellate Body and panels enjoy a wide margin of discretion when interpreting WTO provisions and are not obliged to follow the interpretations advanced by the parties to the dispute.¹² This may be labelled as an intrusive standard of review and it is conventionally justified by the need to maintain consistency in the interpretation of WTO provisions.¹³

In addition, one may also speak about the standard of review applied by a higher reviewing body to determinations made by a lower body (if a review system is based on two instances). In the context of the WTO, this type of review determines the extent of scrutiny of the Appellate Body of a panel's findings (both legal and factual). According to Article 17.6 of the Dispute Settlement Understanding (DSU),¹⁴ such a review is limited to issues of law covered in a panel report and legal interpretations developed by a panel. This in principle excludes any review of factual determinations made by a panel. However, the standard of review applied by a panel when examining

¹² Oesch 2003, p. 18.

¹³ Ehlermann and Lockhart 2004, p. 498. In particular, Ehlermann and Lockhart noted that deferring to the WTO Members' interpretations would lead to the 'Tower of Legal Babel'; this would mean that 'the obligations assumed by WTO Members, and the rights acquired, would differ from Member to Member, undermining the core objectives of the rule-based system'.

¹⁴ Understanding on Rules and Procedures Governing the Settlement of Disputes, 1869 UNTS 401, signed on 15 April 1994.

evidence put forward by the parties to a dispute falls within the purview of the Appellate Body's review. The same is true for 'consistency or inconsistency of a given fact or set of facts with the requirements of a given treaty provision', which is clearly a legal issue.¹⁵

In principle, WTO law does not provide any explicit standard of review to be followed by panels when evaluating and assessing the factual elements of a dispute. The only exception is the Anti-Dumping Agreement,¹⁶ which stipulates in Article 17.6 (i) that:

in its assessment of the facts of the matter, the panel shall determine whether the authorities' establishment of the facts was proper and whether their evaluation of those facts was unbiased and objective. If the establishment of the facts was proper and the evaluation was unbiased and objective, even though the panel might have reached a different conclusion, the evaluation shall not be overturned.

The above standard is conventionally described as a deferential one as it mainly concentrates on procedural rather than substantive compliance (i.e. unbiased and objective evaluation). Consequently, it gives a considerable margin of discretion to national authorities when making factual determinations during the course of an anti-dumping proceeding.

None of the other WTO agreements, including the SPS Agreement, contain any comparable provision. As a consequence, it was for the WTO dispute settlement bodies to identify the applicable standard of review (either as a general standard to be

¹⁵ Appellate Body, *EC – Hormones*, para. 132.

¹⁶ Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994, 1868 UNTS 201, signed on 15 April 1994.

applied across different agreements or a specific variation that would be applicable in the context of a particular agreement). At least on its face, the Appellate Body opted for the first option and identified Article 11 of the DSU as a rule determining the applicable standard of review for the entire WTO system (except for the Anti-Dumping Agreement). According to the Appellate Body in *EC-Hormones*, Article 11 ‘articulates with great succinctness but with sufficient clarity the appropriate standard of review for panels in respect of both the ascertainment of facts and the legal characterization of such facts under the relevant agreement.’¹⁷ On that basis, the Appellate Body vaguely characterized the applicable standard as ‘neither *de novo* review as such, nor “total deference”, but rather the “objective assessment of the facts”.’¹⁸ The Appellate Body also added that:

many panels have in the past refused to undertake *de novo* review, wisely, since under current practice and systems, they are in any case poorly suited to engage in such a review. On the other hand, ‘total deference to the findings of the national authorities’, it has been well said, could not ensure an ‘objective assessment’ as foreseen by Article 11 of the DSU.¹⁹

A closer look at the statement of the Appellate Body reveals, however, its deficiencies. The Appellate Body, by merely describing an applicable standard of review as objective, missed the opportunity to provide future panels with more precise interpretative guidelines. In particular, it was noted that ‘this broad formulation does not assist in defining an operable standard of review because any assessment of the facts, whether highly deferential, marginally deferential, or not deferential at all, can

¹⁷ Appellate Body, *EC – Hormones*, para. 116.

¹⁸ *Ibidem*, para. 117.

¹⁹ *Ibidem*.

be “objective”.²⁰ Although, this observation is probably overstated (at the end of the day the Appellate Body at least identified what types of review are excluded in WTO law), it highlights the vagueness inherent in the statement. Even if one eliminates the extremes (*de novo* review and total deference) the remaining range of options remains quite broad.

One may also have some doubts about the choice of Article 11 of the DSU as the appropriate legal basis for determining the applicable standard of review.²¹ The expression ‘objective’ seems to be more concerned with guarantees, in the context of WTO law, of due process rights (i.e. fairness and impartiality of a panel, neutrality in assessment of presented evidence) rather than with determination of the applicable standard of review.²² In fact, the Appellate Body expressly recognized this aspect of Article 11, when it held that:

[t]he duty to make an objective assessment of the facts is, among other things, an obligation to consider the evidence presented to a panel and to make factual findings on the basis of that evidence. The deliberate disregard of, or refusal to consider, the evidence submitted to a panel is incompatible with a panel’s duty to make an objective assessment of the facts.²³

Nevertheless, the holding of the Appellate Body in *EC - Hormones* has become a point of reference for WTO dispute settlement bodies and since then Article 11 has

²⁰ Bohanes and Lockhart 2009, p. 389.

²¹ Croley and Jackson once suggested that Article 3.2 DSU would be a more appropriate provision. In the relevant part it provides: ‘[r]ecommendations and rulings of the DSB cannot add to or diminish the rights and obligations provided in the covered agreements’ (Croley and Jackson, 1996, p. 199).

²² Guzman 2008, p. 4

²³ Appellate Body, *EC – Hormones*, para. 133.

been cited as a rule that elucidates the required level of scrutiny. Thus, Article 11 may be regarded as having a dual function: a provision that determines (very imperfectly) the applicable standard of review, and a rule which establishes due process rights for the parties to a dispute.

Although the objective standard of review was introduced as a general rule applicable across various WTO agreements, a substantive analysis of the various panel reports shows that in practice different agreements attract different types of review.²⁴ As discussed in section 3.3 below, this is also true for various requirements contained in one agreement.

3. Standard of review and the SPS Agreement

3.1. SPS Agreement and its disciplines

The SPS Agreement is particularly important for the settlement of health and environment-related trade disputes. The major aim of agreement is to limit the impact on international trade²⁵ of national SPS measures (i.e. measures which aim at protection of human, animal and plant life and health against some specifically enumerated risks).²⁶ At the same time, the agreement intends to guarantee to WTO Members a wide margin of regulatory discretion in the SPS area. Consequently, while WTO Members are expected to observe certain requirements when introducing and

²⁴ Ehlermann and Lockhart 2004, pp. 503-521.

²⁵ As noted in the literature, domestic health and environmental measures ‘can often be manipulated or exploited to protect domestic industry from international competition’ (Trebilcock and Soloway, 2002, p. 537).

²⁶ The relevant risks include risks arising from additives, contaminants, toxins or disease-causing organisms in foods, beverages or feedstuffs. On more detailed discussion on the applicability of the SPS Agreement see Chapter [] of this book.

implementing their SPS measures, they remain in principle free to establish whatever level of protection they deem appropriate. In WTO nomenclature this right is conventionally referred to as the right to establish the appropriate level of protection (ALOP) and indicates the maximum SPS risk that a particular WTO Member is ready to tolerate.

This discretion is reflected in Article 3 of the SPS Agreement. As a general rule WTO Members are obliged to base their SPS measures on international standards, guidelines, and recommendations (Article 3.1),²⁷ however under certain conditions they may also deviate therefrom (Article 3.3). This option becomes available if there is sufficient scientific evidence to support a domestic measure, i.e. a deviating measure needs to be based on scientific principles and cannot be maintained without sufficient scientific evidence.²⁸ This general instruction is translated into the specific requirement of scientific risk assessment. Thus, Article 5.1 stipulates that Members have to ensure that their SPS measures are based on an assessment, as appropriate to the circumstances, of the risks to human, animal or plant life health. The agreement also enumerates elements that need to be included in such assessment (Articles 5.2–5.3). These cover not only available scientific evidence, but also other relevant elements such as processes and production methods, ecological and environmental conditions, and economic factors (but only for quarantine risks). WTO Members may also act if the available scientific evidence is insufficient to perform an adequate risk assessment. A provisional measure, based on available pertinent information, is then a

²⁷ The category of relevant international organizations includes: the Codex Alimentarius Commission, the World Organization for Animal Health and the International Plant Protection Convention. SPS measures, which conform to such international standards, are presumed to be in compliance with the SPS Agreement.

²⁸ Article 2.2 of the SPS Agreement.

permissible option. In such a case, a Member should seek to obtain the additional information necessary for more objective assessment of risk and review of its SPS measure within a reasonable period of time (Article 5.7).

Therefore, under the SPS Agreement, science operates as a criterion which allows one to distinguish between permissible and prohibited measures. It may be seen either as a proxy (although imperfect) for detection of protectionist measures taken in the guise of health and environmental regulations, or as a method of improving market access by introducing a certain technical rationality (at the end of the day a measure is condemned irrespective of whether it has a protectionist character or not).²⁹ Such a mechanism requires panels, as sole fact-finders in the WTO dispute settlement process, to assess and evaluate scientific claims made by the parties to the dispute. This obviously raises the question of the appropriate standard of review over scientific determinations made on a national level.

A separate set of SPS obligations is imposed on the risk management phase of the national regulatory process. As noted above, WTO Members may in principle adopt any appropriate level of protection (ALOP). At least in theory this also encompasses a zero risk level, even if potential costs to international trade clearly exceed expected health/environmental benefits.³⁰ On the other hand, the Agreement introduces in Article 5.5 the idea of consistency in ALOP for domestic risk-related regulations and requires a certain level of uniformity in different but comparable risk situations (e.g. regulatory response with respect to the same or similar pathogen or disease). This

²⁹ On the role of science in the SPS Agreement see Gruszczynski, 2010, pp. 147-155 and literature cited there.

³⁰ The Agreement imposes in this regard a soft obligation that only requires Members, when determining the appropriate level of SPS protection, to take into account the objective of minimizing negative trade effects (Article 5.3).

obligation is, however, not absolute as Members may still differentiate in their regulatory reactions if they are able to provide persuasive justification or show that there is no arbitrariness. Members also need to ensure that their SPS measures are not more trade-restrictive than necessary to achieve their ALOP, taking into account technical and economic feasibility (Article 5.6). In this context, the SPS Agreement identifies two concepts that may help in ensuring least-trade restrictiveness: regionalisation (i.e. adapting domestic SPS measures to the specific conditions prevailing in the place of the origin and import destination) and equivalence (i.e. accepting measures of other Members as equivalent to domestic ones, if they guarantee the same level of protection). In addition, measures have to be applied only to the extent necessary to protect human, animal or plant life or health (Article 2.2) and cannot arbitrarily or unjustifiably discriminate between Members where identical or similar conditions prevail, including between the territory of the regulating Member and that of other Members (Article 2.3).

3.2. Standard of review under the SPS Agreement – initial developments

The SPS Agreement is silent about applicable standard of review. In the first SPS case (*EC – Hormones*), the EC argued on appeal that the panel was obliged to apply a deferential reasonableness standard, or the standard of review that is provided by Article 17.6(i) of the Anti-Dumping Agreement. According to the EC, such a standard would be applicable to ‘all highly complex factual situations, including the assessment of the risks to human health arising from toxins and contaminants.’³¹ In practical terms, this would mean a concentration on procedural rather than substantive compliance (i.e. whether a procedure prescribed by SPS Agreement was followed,

³¹ Appellate Body Report, *EC – Hormones*, para. 112.

without going into the substance of a national SPS measure).³² The Appellate Body disagreed. It found no indication in the SPS Agreement that Members wanted to incorporate the standard set out in Article 17.6(i) of the Anti- Dumping Agreement. Instead, as was already mentioned, the Appellate Body identified Article 11 of the DSU as providing the applicable standard of review.

Article 11 calls for an objective standard of review which, according to the Appellate Body, is neither *de novo* review nor total deference. The analysis of the *EC – Hormones* report shows that in practice the Appellate Body opted for a rather deferential approach.³³ Under this standard, a panel, although entitled to examine the underlying science and scientific evidence, has to grant a WTO Member a relatively broad degree of deference. In particular, the Appellate Body recognized that Members are entitled to base their SPS measure not only on the mainstream (i.e. the best available) science, but could also rely on minority scientific opinions.³⁴ This obviously limits the discretion of panels with regard to the re-assessment of scientific evidence. As a consequence, a panel is not allowed to condemn a national measure based on a finding that a majority of scientists hold a different view from the one supporting a domestic measure. Second, risks that need to be evaluated under Article 5.1 are not only those which are ‘ascertainable in a science laboratory operating under strictly controlled conditions’ but also real world risks that take into account enforcement problems, human errors etc.³⁵ Again this gives Members an opportunity to take into account different factors and to contextualize risk in a specific national

³² EC appellant submission in *EC - Hormones*, para. 126.

³³ Cf. Thomas 1999, p. 507, Bloche 2002, p. 837.

³⁴ Appellate Body Report, *EC – Hormones*, para. 194.

³⁵ *Ibidem*, para. 187.

setting. Third, the required connection between risk assessment (scientific evidence) and the SPS measure was characterized as merely a reasonable one, as opposed to the more demanding standard of strict conformity. Fourth, the Appellate Body instructed the panel to ‘bear in mind that responsible, representative governments commonly act from perspectives of prudence and precaution where risks of irreversible ... damage to human health are concerned.’³⁶ This implies that an additional margin of discretion needs to be granted to national governments when evaluating those measures aimed at eliminating irreversible risks to human health and life.

The subsequent case law, however, gradually engaged in a more and more intrusive assessment of the scientific data provided as a justification for domestic SPS measures. Such a standard of review allowed WTO panels to assess the quality, persuasive force, and correctness of scientific determinations made on national levels and to substitute them with their own. In practice this came very close to *de novo* review. In *Australia – Salmon*, the Appellate Body made clear that the panel was not required to ‘accord to factual evidence of the parties the same meaning and weight as do the parties.’³⁷ In a similar fashion in *Japan – Apples* the Appellate Body explained that the panel did not need to favour Japan’s approach to risk and scientific evidence over the views of its own experts.³⁸ Consequently, the panel had a considerable margin of discretion in assessing the value of the evidence and the weight to be ascribed to such evidence. As explained by the Appellate Body ‘requiring panels (...) to give precedence to the importing Member’s evaluation of scientific evidence and

³⁶ *Ibidem*, para. 181.

³⁷ Appellate Body Report, *Australia – Measures Affecting Importation of Salmon*, WT/DS18/AB/R, adopted 6 November 1998, DSR 1998:VIII, 3327, para. 267.

³⁸ Appellate Body Report, *Japan – Measures Affecting the Importation of Apples*, WT/DS245/AB/R, adopted 10 December 2003, DSR 2003:IX, 4391, para. 165.

risk is not compatible with this well-established principle [of objective review].³⁹ Indeed, the analysis of the report in *Japan – Apples* shows that the panel did replace the factual determinations of national authorities (i.e. Japanese) with its own.⁴⁰ On the other hand, one may still find traces of the earlier more deferential approach. The panels consistently emphasised that they could not engage in *de novo* review of scientific determinations made by WTO Members. For example, the panel in *Japan – Agricultural Products* explained that its duty was not to conduct a new risk assessment, but rather to assess the quality of the national evaluation.⁴¹ The SPS case law, at least on its face, also confirmed that Members could base their measures on minority scientific opinions.

The idea of engaging in a rather aggressive examination of scientific evidence was followed and developed further by the panel in *EC – Biotech Products* case.⁴² The panel conducted a detailed and intrusive analysis of the scientific evidence that was put forward as a justification for the EC measures. It chose between competing scientific claims articulated by its experts, preferring some over others. The panel's analysis of the scientific materials submitted by the EC under Article 5.7 may serve as a good example of this approach. The aim of panel's examination was to determine whether the evidence was insufficient, and therefore justified the defendant's recourse to the disciplines of Article 5.7. In this context, the panel was required to evaluate a

³⁹ *Ibidem*, para. 167.

⁴⁰ *Cf.* Prevost 2005, p. 11.

⁴¹ Panel Report, *Japan – Measures Affecting Agricultural Products*, WT/DS76/R, adopted 19 March 1999, as modified by Appellate Body Report WT/DS76/AB/R, DSR 1999:I, 315, para. 8.32.

⁴² Panel Report, *European Communities – Measures Affecting the Approval and Marketing of Biotech Products*, WT/DS291/R, WT/DS292/R, WT/DS293/R, Add.1 to Add.9, and Corr.1, adopted 21 November 2006, DSR 2006:III-VIII, 847.

range of scientific data relied on by the competent authorities in several EU Member States (e.g. the opinion of the French Biomolecular Engineering Committee) as well as more recent scientific materials. The panel also consulted six independent experts on some specific scientific issues. In line with the *Japan – Apples* ruling, the panel did not see its task as limited (i.e. to whether evidence relied on by the EC reasonably supported the conclusion of the existence of insufficiency) but rather adopted the role of final arbiter. Thus when assessing the evidence on oilseed rape MS1xRF1, it arbitrarily disregarded not only the view expressed by the Biomolecular Engineering Committee, but also the opinions of some of its own experts who seemed to support the position of the EC.⁴³ What the panel considered important was the fact that the Scientific Committee on Plants (a body responsible for evaluation of risk at the EU level) was able to perform relevant risk assessment. The panel did not explain why it did not pay attention to other evidence.⁴⁴

The panel in the *US/Canada – Continued Suspension* disputes took essentially the same approach.⁴⁵ It confirmed that it enjoyed a broad discretion as to the choice of evidence when making factual findings. It also added that it was not ‘expected to refer to all statements made by [its] experts and should be allowed a substantial margin of discretion as to which statements are useful to refer to explicitly as long as [it does]

⁴³ See generally, Gruszczynski 2010, pp. 194-195.

⁴⁴ Panel Report, *EC – Biotech Products*, para. 7.3300.

⁴⁵ Panel Report, *United States – Continued Suspension of Obligations in the EC – Hormones Dispute*, WT/DS320/R, adopted 14 November 2008, as modified by Appellate Body Report WT/DS320/AB/R and Panel Report, *Canada – Continued Suspension of Obligations in the EC – Hormones Dispute*, WT/DS321/R, adopted 14 November 2008, as modified by Appellate Body Report WT/DS321/AB/R. I refer only to the *US – Continued Suspension* case hereafter, as the findings in both reports are almost identical.

not deliberately disregard or distort evidence.’⁴⁶ The panel admitted that under the SPS Agreement, it was not asked to carry out its own risk assessment, but it also explained that in fact its position was similar to that of a national body producing such an assessment. This meant that the panel wanted to receive the full spectrum of expert views to form an opinion as to the correctness of the risk assessment.⁴⁷ Although the *Suspension* panel confirmed that a measure could be based on minority scientific opinions, in practice it adopted the role of an ultimate arbiter as to what could be considered scientific and what not. The following passage from the panel report precisely summarises its approach:

while, on some occasions, we followed the majority of experts expressing concurrent views, in some others the divergence of views were such that we could not follow that approach and decided to accept the position(s) which appeared, in our view, to be the most specific in relation to the question at issue and to be best supported by arguments and evidence.⁴⁸

Overall, this approach may be described as a ‘*quasi de novo*’ standard of review.

The direction of the SPS case law, which favours a rather intrusive inquiry into national scientific determinations, is problematic in many respects. From the pragmatic point of view, panels seem to lack sufficient scientific competence to make a judgement over complex technical and scientific issues. Although they are assisted by experts, it is still ‘very difficult for them to be sure that they are focusing on the most relevant statements’⁴⁹ or that they correctly appreciate the value and relevance

⁴⁶ *Ibidem*, para. 7.416.

⁴⁷ *Ibidem*, para. 7.418.

⁴⁸ *Ibidem*, para. 7.420.

⁴⁹ Epps 2011, p. 16

of specific scientific claims. Panellists also need to decide how to interpret the opinions of their experts, how to assess and chose between conflicting views, or what kind of inferences to make therefrom. This is rarely an easy task. Note that the answers of experts are frequently formulated in conditional language to reflect the uncertainties inherent in scientific research. This qualified character of the scientific discourse relating to identification and assessment of health and environmental risks is well captured by the group of eminent scholars who noted that:

the complete description of particular risks usually looks as follows:
'exposure to [X] carries with it a probability of adverse effect that varies from person to person, but is generally in the range of (for example) one in a million to ten in a million; this range is subject to uncertainty and so the true range may be from one in a million to a hundred in a million or may be even zero.'⁵⁰

Unless a panel grants the WTO Members a considerable degree of discretion, there is the risk that some legitimate measures will be condemned not because they lack scientific basis, but because of subjective differences in assessment of the scientific evidence between a particular national government and a panel. This, in turn, will bring the right of WTO Members to adopt SPS measures that achieve their ALOP into question.

Even if there is no qualification, replies may fail to provide straightforward yes or no answers. In such a case, a panel will either need to conduct an additional investigation into a particular scientific problem, or simply decide the issue on the basis of the received answers (e.g. by choosing those answers which appear to it as more direct or pertinent). This option may be tempting for a panel – at least on its face

⁵⁰ Crawford-Brown, Pauwelyn and Smith 2004, p. 463.

it makes the task of a panel easier and speeds up the whole fact-finding process. However, it also creates a risk of misunderstanding and simplification, or at least some arbitrariness in selecting evidence. The discussion on genotoxicity in *US/Canada – Continued Suspension* illustrates this problem. There was a fundamental disagreement between the parties to the dispute as to the existence of a threshold for substances having genotoxic potential (in this case for oestradiol-17 β). The EC claimed that no threshold could be identified, meaning that there was no level below which intakes from hormone residues should have been considered safe. This also meant that the size of doses used in the growth promotion process was of no relevance (at least as far as the existence of risk was concerned, not its extent). Both Canada and the US maintained that the threshold for oestradiol-17 β could be determined. In this context, the panel asked its experts a very specific question: whether the scientific evidence referred to by the EC supported its position. The answer of Dr Boisseau was straightforward. He confirmed that the scientific evidence referred to by the EC did not demonstrate a no-threshold mechanism for oestradiol-17 β .⁵¹ Dr Cogliano was of different opinion and explained that ‘the EC’s statement that a threshold cannot be identified reflects their view of genotoxic mechanisms, just as the contrary statement that there is a threshold and that this threshold is above the levels found in meat residues reflects how Canada and the US view genotoxic mechanisms.’⁵² He also clarified that none of the statements was scientifically demonstrated, and that both positions could be considered as simply based on different assumptions used in the

⁵¹ Panel Report, *US – Continued Suspension, Annex D, ‘Replies of the Scientific Experts to Questions Posed by the Panel’*, para. 182.

⁵² *Ibidem*, para. 186.

interpretation of the available evidence.⁵³ Dr Guttenplan, the third expert answering this question, was the most ambiguous. He noted at the beginning that there was no reason to expect a threshold to exist for a genotoxic chemical. However, he explained that the statement that ‘the fact that doses used in growth promotion are low is not of relevance [was] not necessarily true.’⁵⁴ In this context, he clarified that a dose always determined the risk (i.e. low exposure produces low risk). Finally, he added that ‘at very low levels of genotoxic carcinogens the decrease in risk [was] more than proportional than the decrease in applied dose.’⁵⁵ On the basis of the above answers one may ask what is the relevance of the observation made by Dr Cogliano? To what extent did the EU's reliance on a non-scientific assumption affect its overall position? What does the expression ‘not necessarily true’ used by Dr Guttenplan mean? Does it indicate a genuine disagreement between scientists, or it is simply an expression of diligence and precaution in making definitive scientific statements? Unfortunately, the report does not provide the answers to these questions. The panel apparently accepted the answer of Dr Boisseau as being the most straightforward (and therefore correct) and simply disregarded the others. As a consequence, it found that the EC did not submit evidence which would indicate that there is no threshold for oestradiol-17 β .⁵⁶

The risk of failing to properly appreciate the scientific value and importance of specific claims and evidence is aggravated by the institutional infrastructure that the panels rely on and the time constraints built into the dispute settlement system. The

⁵³ *Ibidem*.

⁵⁴ *Ibidem*, para. 187.

⁵⁵ *Ibidem*.

⁵⁶ Panel Report, *US – Continued Suspension*, para. 7.572.

WTO is an international organization which possesses specialized knowledge (and one arguably superior to that of its Members) in matters relating to international trade, but not with respect to scientific issues relating to protection of health and life of humans, animals and plants. To put it differently, the WTO lacks the capacity to ‘undertake its own inquiry into the science.’⁵⁷ If one combines this fact with tight time constraints imposed on panels by the DSU, an in-depth examination of the scientific evidence put forward by the parties hardly seems to be an advisable option.

Lack of expertise is, however, not the only problem faced by the dispute settlement bodies when evaluating scientific evidence. As recognized in the literature, assessment of risk is not a purely scientific task, and depends on the specific socio-cultural conditions of a particular country.⁵⁸ Such assessment inevitably involves the subjective judgments of assessors and reflects their attitudes toward particular risks, values of the relevant community in which the experts are acting, and other normative elements such as required level of protection or approach to uncertainties. Consequently, the same set of scientific data may produce entirely different risk estimates in various jurisdictions. The panel seems to be worse placed to make such judgements than the WTO Member that conducted research and evaluated scientific evidence in the context of its specific risk frames and concerns.⁵⁹ If a panel ends up imposing on the WTO Members its own vision of science, these normative and context-dependent elements will be lost, and the ultimate determination may fail to produce a correct result. Although, the *de novo* standard of review allows for avoiding what are conventionally referred to as Type 2 errors (i.e. allowing protectionist or

⁵⁷ Guzman 2007, p. 229

⁵⁸ See generally, Winickoff *et al* 2005.

⁵⁹ *Cf.*, Button 2004, p. 181.

‘unnecessary’ measures to escape scrutiny from WTO obligations), it may also produce so-called Type 1 errors (i.e. condemning a measure that actually protects health and safety). Is it worth losing one statistical life to achieve some additional trade liberalization? How does one compare the damage to environment caused by the invasion of foreign species with the potential gains generated by an increase in trade? These are difficult decisions that should be addressed rather at the national than international level.

These questions bring us to yet another issue, which seems to be particularly important in the context of health and environment-related trade disputes. The potential costs of a mistake on the side of the WTO dispute settlement bodies appear to be considerably higher in SPS disputes, as compared to other trade disagreements. A loss of statistical life as a consequence of removing a trade barrier is arguably more costly (as a matter of principle and not in terms of specific monetary value) than any damage to international trade. Some additional costs may also be generated by problems in compliance. In particular, it is noted that national health and safety measures regulate very sensitive areas which were always considered within the core of national sovereignty. Due to the high values at stake, it may be simply politically impossible (arguably more frequently than under other WTO agreements) to comply with a WTO ruling. Potential costs here relate not only to additional distortions of international trade resulting from the suspension of a concession in response to non-compliance, but also to the harm that is caused to the credibility of the WTO as such. If Members do not comply with rulings, the whole dispute settlement system is called into question.⁶⁰

Another factor which speaks against *de novo* review is the impact that such a

⁶⁰ Guzman 2007, p. 230.

standard has on the length of the dispute settlement process. Note that one of the functions of the WTO is the efficient settlement of disputes. In particular, Article 3.3.

DSU provides that:

the prompt settlement of situations in which a Member considers that any benefits accruing to it (...) under the covered agreements are being impaired by measures taken by another Member is essential to the effective functioning of the WTO and the maintenance of a proper balance between the rights and obligations of Members.

Elaborating on this rule, Article 12.9 DSU stipulates that as a general rule a panel is required to issue its report within six months. An extension is possible but it should not exceed nine months.⁶¹ *De novo* review, which requires a detailed examination of complex scientific evidence, prolongs that process considerably as compared to disputes decided under the other agreements. The *EC - Biotech* and *US/Canada – Continued Suspension* cases are good examples here. Both of them lasted for years (3 and almost 4 respectively). The same is true for the latest *Australia – Apples* dispute, where the publication of the panel report was postponed several times and the whole procedure continued for almost 3 years. One may legitimately ask whether such period of time meets the requirement of prompt settlement of a dispute provided for by the DSU. A more deferential standard under which a panel would concentrate on methodological rather than substantive issues has the potential to considerably shorten that process.

The above concerns provoked a group of scholars to argue that WTO panels should not function as adjudicatory bodies reviewing the substantive scientific details of domestic risk assessments, but rather as administrative bodies that only supervise

⁶¹ In practice a panel proceeding takes on average about 12 months.

transparency and procedural aspects of the national regulatory process.⁶² Another author proposed the limitation of a panel's review to the assessment whether a measure is arbitrary/unjustifiable discriminatory, constitutes the least trade restrictive alternative, or complies with the transparency requirements of Annex B of the SPS Agreement. According to him, any scrutiny into the scientific value of a measure, national preferences for risk or the relationship between risk assessment and a measure, should be very restricted, with a panel concentrating on procedural rather than substantive issues (whether a government took scientific evidence into account, not whether such evidence supports a measure under the examination).⁶³ This would amount to very deferential standard of review.

The problem with such proposals is that they seem go against the text of the SPS Agreement, which requires some form of substantive scrutiny of both risk assessment and scientific evidence. Article 3.3 allows deviation from international standards when there is scientific justification (and not merely when a Member complies with some procedural rules). Both Article 2.2 and 5.1 require a scientific basis for an SPS measure. This implies at least some kind of enquiry (more or less deferential) into the substance of evidence supporting a measure. This language arguably cannot be reduced to a simple procedural review. Similarly, the definition of risk assessment relies on substantive rather than procedural factors (e.g. it refers to examination of probability/possibility and not to the risk assessment process as such). This is also true for other provisions of the SPS Agreement, which in most cases are of a substantive character - e.g. Article 5.7 and its obligation to base a measure on

⁶² Winickoff *et al* 2004, p. 109.

⁶³ Guzman 2007, p. 231.

pertinent information.⁶⁴ The procedural requirements of the SPS Agreement remain general and relatively underdeveloped. Consequently, they can hardly be regarded as benchmarks against which national measures can be effectively tested and assessed. Moreover, without denying the advantages of procedural checks (i.e. increased transparency of decision-making processes, which may help to detect instances when a measure is adopted due to pressure from rent-seeking groups), such a model will be most probably under-inclusive, with many measures escaping the scrutiny of WTO dispute settlement bodies. As noted by Trebilcock and Soloway ‘if too wide a degree of deference is afforded to the Member’s regulation, and any remotely plausible explanation can be offered as a rationale for trade-restrictive health and safety standards, the world trading system risks being seriously undermined with attendant global and domestic welfare losses in gains from trade.’⁶⁵ This might suggest that although adopting a fully deferential standard of review in the context of the SPS Agreement is not an advisable option, a considerable degree of deference is still called for.

3.3. Standard of review under the SPS Agreement – some recent developments

The 2008 Appellate Body’s report in the *US/Canada – Continued Suspension* cases may be seen as a response to the above concerns. At the time of its adoption, it appeared to be a revolutionary shift in WTO jurisprudence, with the Appellate Body opting for a more deferential approach which would grant an additional margin of discretion to national governments in the SPS area.

One of the issues contested by the EC in the appeal was the standard of review

⁶⁴ Cf. Goh 2006, pp. 667-668.

⁶⁵ Trebilcock and Soloway 2002, p. 541.

applied by the panel to factual determinations, including scientific evidence. The Appellate Body first observed, rather uncontroversially, that it was the task of each Member State to perform a risk assessment, while a panel was only expected to review it. This meant to the Appellate Body that it was not for the panel to determine whether a risk assessment was correct but only whether it was supported by coherent reasoning and respectable scientific evidence.⁶⁶ The Appellate Body went on to identify four specific steps that should have been taken by the panel when performing this limited task. Thus, a panel must:

- a) identify the scientific basis underlying an SPS measure;
- b) verify that the scientific basis comes from a respected and qualified source - this should be relatively easy for mainstream scientific claims, whereas for minority scientific opinions some more detailed examination will be required;
- c) assess whether the reasoning articulated on the basis of the scientific evidence is objective and coherent - in other words, the panel needs to assess whether a conclusion reached by a WTO Member finds sufficient support in scientific evidence relied upon; and
- d) determine whether the results of the risk assessment sufficiently warrant the SPS measure at issue.⁶⁷

The Appellate Body again confirmed that both mainstream science as well as minority scientific opinions could serve as a basis for a measure. What is important is epistemic value of a particular claim. As explained by the Appellate Body, ‘evidence must have necessary scientific and methodological rigour to be considered reputable

⁶⁶ Appellate Body Report, *US – Continued Suspension*, para. 590.

⁶⁷ *Ibidem*, para. 591.

science'.⁶⁸ This aspect is to be assessed using the methodological parameters of the relevant scientific community. Note again that the question here is not whether in the opinion of a particular scientist the evidence is correct, but rather whether specific information is defensible as a scientific claim.

Once a scientific basis is established and its epistemic status verified, a panel would need to examine the coherence and objectivity in the interpretation of these raw data. Although the Appellate Body did not elaborate on this particular point, one could legitimately expect that a task of a panel would also be limited here. This would mean that conclusions drawn by a WTO Member on the basis of scientific evidence are only to be checked against scientific logic (i.e. whether a specific interference is justified in the light of a particular methodology that may be described as scientific, and whether an interference is objective or biased). However, as will be discussed below, the subsequent case law decided that panel's scrutiny is actually more intrusive when evaluating coherence and objectivity of reasoning (as compared to scientific evidence as such).

Finally, a panel needs to inquire into the relationship between the conclusions of a risk assessment and an SPS measure. This is the second point where the Appellate Body was rather enigmatic. It remains unclear what level of compatibility between those two elements is required. One may assume, in line with the existing case law, that scientific findings may support a whole range of different SPS measures. As noted elsewhere by the panel in *EC - Biotech*:

there may conceivably be cases where a Member (...) would be justified in applying (i) an SPS measure even though another Member might not decide to apply any SPS measure on the basis of the same risk assessment, or (ii) an

⁶⁸ *Ibidem*.

SPS measure which is stricter than the SPS measure applied by another Member to address the same risk.⁶⁹

In any case, some additional guidance for assessing the existence of a sufficient relationship between the conclusions of risk assessment and an SPS measure seems to be desirable.

The new methodology proposed by the Appellate Body also has an impact on the consultation process that takes place between a panel and its experts. In particular, a panel may seek the assistance of experts in order to examine the individual steps described above (e.g. to identify the scientific basis of an SPS measure, determine whether it constitutes a defensible scientific claim, or review whether it is objective and coherent). In this context, the Appellate Body warned that the panel should not attempt to test whether its experts would have done a risk assessment in the same way as a particular WTO Member. Their task is limited, as they should only review an assessment in terms of its scientific value (i.e. whether it is defensible).⁷⁰ This obviously determines the extent of scientific advice that may be sought by a panel - e.g. what questions are appropriate and what kind of answers are to be considered relevant when deciding a particular issue.

The above approach differs considerably from the approach taken in the previous case law. It explicitly prohibits a panel from inquiring into the correctness of evidence and instructs it to concentrate on methodological issues in order to assess epistemic value and the coherence of scientific findings. A panel may equally check (in order to verify the objectivity of a process) whether municipal authorities have collected and considered all relevant evidence and whether the evaluation of this evidence was

⁶⁹ Panel Report, *EC – Biotech Products*, para. 7.3065.

⁷⁰ Appellate Body Report, *US – Continued Suspension*, para. 592.

unbiased. This standard leaves WTO Members with a greater degree of discretion as to how to assess scientific data and what kind of interferences to make on their basis. It also greatly reduces the need for a panel to engage in a detailed examination of scientific evidence and decide which scientific view is better. In theory, once it is established that a particular claim is scientifically defensible, even if improbable, the task of the panel ends.

The initial reactions from academic circles to the report of the Appellate Body were rather positive. Some even claimed that the Appellate Body had decided to adopt a predominantly deferential approach, under which the main focus would be on the process of risk assessment rather than its outcome. In this context, a parallel was drawn between the approach of the Appellate Body and that of the US Supreme Court in the *Chevron* case.⁷¹ Others, however, were more sceptical. For example, Peel, although labelling the approach of the Appellate Body as procedural in nature, also recognized that it 'still contains a substantial emphasis on scientific factors.'⁷² Indeed the tests provided by the Appellate Body appear to go beyond a merely procedural approach that concentrates only on formal compliance. Arguably points (c) and (d) (i.e. whether the reasoning articulated on the basis of the scientific evidence is objective and coherent and whether the results of the risk assessment sufficiently warrant the SPS measure at issue) require the performance of a substantive analysis (albeit deferential one). Moreover, an inquiry into underlying methodology (in order to verify that a specific claim is regarded as reputable science) can be also connected with different levels of scrutiny. As will be elaborated further below, a detailed methodological assessment may in practice come quite close to *de novo* review.

⁷¹ Cf. Robert Howse on the International Economic Law and Policy Blog (<http://worldtradelaw.net>)

⁷² Peel 2010, p. 218.

Nevertheless, the overall impression that one gets is that the Appellate Body ‘open[ed] the door a little wider to recognition of a greater diversity of risk assessment approaches in the SPS context.’⁷³ The holding that the proper task of the panel is not to assess the correctness of domestic scientific determinations but rather their overall coherence and objectivity is important. It suggests a rather deferential approach that considerably differs from the previous practice. Having said that, it is also true that the Appellate Body left many important interpretative questions open.

Australia – Apples was the first case to elaborate on this new approach. The applicable standard of review actually turned out to be one of the major issues in the appeal lodged by Australia. Although the Appellate Body provided rather extensive legal analysis, the report leaves a reader disappointed and perplexed. Its tone is very different from the spirit of the *US/Canada – Continued Suspension* report and indicates that the investigation of the WTO panel remains relatively intrusive.

The dispute arose in the context of an Australian import prohibition applicable to New Zealand apples on phytosanitary grounds. At least in theory, Australia was concerned with the risks posed by two pests (i.e. European canker and apple leafcurling midge) and one disease (i.e. fire blight), which were absent on its territory but had been reported for New Zealand. The Australian risk assessment, which was completed in 2006, recognized New Zealand apples as a potential transmission vector. It listed a number of specific conditions that had to be met before any export could take place. New Zealand contested the measure in the WTO dispute settlement system and the panel found in its favour. On appeal, Australia contested number of different issues, including the standard of review that was applied by the panel in its assessment of scientific determinations. In essence, Australia argued that the panel,

⁷³ *Ibidem*.

following the instruction of the Appellate Body in *US/Canada – Continued Suspension*, was obliged to adopt a relatively deferential approach.

The Appellate Body, however, rejected the Australia's assertion that the panel had exceeded its mandate. The Appellate Body started its analysis with its traditional statement that the applicable standard of review was neither *de novo* review nor full deference. It also recalled that the panel was expected to review a contested risk assessment and not to carry out such assessment on its own.⁷⁴ Interestingly, the Appellate Body identified two different standards of review to be applied in the context of Article 5.1 (and arguably also under Article 2.2). The first one relates to the evaluation of the scientific basis as such (steps (a) and (b) of the *Continued Suspension* test) and is rather deferential. This was correctly justified by the epistemic superiority of a domestic risk assessor ('a panel is not well suited to conduct scientific research and assessments itself and should not substitute its judgement for that of a risk assessor'⁷⁵).

The second type is concerned with the review of reasoning articulated on the basis of scientific evidence, which is a step (c) in the *Continued Suspension* test (i.e. whether the scientific evidence supports the conclusions of the risk assessment to a sufficient degree).⁷⁶ Although the Appellate Body did not identify the applicable standard of review for deciding whether the results of the risk assessment sufficiently warrant the SPS measure (step (d)), one may expect the same level of scrutiny here. Contrary to the first type, this standard of review involves a relatively intrusive examination.

⁷⁴ Appellate Body Report, *Australia – Apples*, paras. 212-13.

⁷⁵ *Ibidem*, para. 225.

⁷⁶ *Ibidem*, para. 215.

On the basis of this distinction, the Appellate Body rejected Australia's claim that the panel should have only evaluated whether 'intermediate conclusions were "within a range that could be considered legitimate" according to the standards of the scientific community.'⁷⁷ Since conclusions (both immediate and ultimate) constitute a part of reasoning and they are subject to more intense scrutiny. The Appellate Body also disagreed with Australia that the panel's review should be limited to the ultimate conclusion reached in a municipal risk assessment. The mandate of the panel covered examination of the reasoning, which also included intermediate conclusions in a risk assessment. The same was true for expert judgements used in a risk assessment to compensate for missing scientific data or to address scientific uncertainties. In this context, the Appellate Body concurred with the panel that Australia was obliged to show that the exercise of such expert judgments was sufficiently documented, transparent, and based on the relevant reliable scientific information.⁷⁸ The Appellate Body believed that the documentation and transparency requirement was 'instrumental in the determination of whether the overall risk assessment, even when it is conducted in the face of some scientific uncertainty, relies on the available scientific evidence.'⁷⁹ Nor did the Appellate Body accept Australia's assertion that the panel's role was limited to determining whether alleged flaws in the reasoning of risk assessment were sufficiently serious to undermine 'reasonable confidence' in the assessment as a whole. This threshold was considered to be too low.⁸⁰

As noted at the outset, the overall approach of the Appellate Body to the

⁷⁷ *Ibidem*, para. 231.

⁷⁸ *Ibidem*, para. 248.

⁷⁹ *Ibidem*, para. 244.

⁸⁰ *Ibidem*, paras. 259-60.

applicable standard of review remains ambiguous. On one hand, some elements in its analysis indicate that it opted for quasi-deferential standard that follows the test articulated in the *Continued Suspension* report. The Appellate Body stressed that the panel's role is limited when 'reviewing whether the scientific basis constitutes legitimate science'⁸¹ (steps (a) and (b) of the test). A panel is only expected to determine that a specific claim can be regarded as 'legitimate' science (i.e. whether it holds a minimum epistemic value and not whether it is correct). It also accepted the panel's approach to concentrate on the methodology used in Australia's risk assessment rather than its ultimate outcome. This is a characteristic feature of deferential types of review.⁸² Similarly, the documentation and transparency requirements (i.e. how risk assessors reached specific expert judgments), which were introduced by the panel and subsequently upheld by the Appellate Body, may be also regarded as a tool for deferential review, as it concentrates on the process of risk assessment and not its substance.

On the other hand, the Appellate Body rejected Australia's argument that the panel's task, when reviewing intermediate conclusions in its risk assessment, should consist in deciding whether they 'fall within a range that could be considered legitimate by the scientific community'. One may also assume that the same approach would be taken for expert judgments, which also constitute a part of reasoning articulated on the basis of scientific evidence. Although the report did not specify the precise level of this more intrusive review, its finding may be seen as an invitation for future panels to assess the correctness of such reasoning rather than its reasonableness. It is not clear why the Appellate Body decided that a separate and

⁸¹ *Ibidem*, para. 215.

⁸² *Ibidem*, paras. 258-259.

more demanding standard of review was applicable to the reasoning of risk assessors. In support of its conclusion, the Appellate Body referred to the *Continued Suspension* report and explained that in this case it ‘considered that the manner of scrutinizing the underlying scientific evidence differs from the manner of scrutinizing the reasoning of the risk assessor’. However, an examination of the relevant paragraphs shows that no such a distinction was made in the *Continued Suspension* case. To the contrary, the deferential standard of review identified there appeared to be a uniform concept that is applicable to different steps of the risk assessment process. What seem to be even more problematic are the reasons which could justify such a distinction. The Appellate Body, when introducing a deferential standard of review for scientific evidence, was primarily concerned with the limited epistemic competence of a panel in scientific matters. However, a panel is in no way better equipped to decide on the correctness of intermediate conclusions in a risk assessment or expert judgments used to compensate for uncertainties (or more generally – the reasoning in a risk assessment). Both aspects are highly complex and require the involvement of experts to advise a panel. In this sense, the challenge that is faced by a panel when evaluating reasoning is not so different from that of scrutinizing scientific evidence as such. This fact was indirectly recognized by the Appellate Body in *US/Canada – Continued Suspension*, when it observed:

The panel may seek the experts’ assistance in order to identify the scientific basis of the SPS measure and to verify that this scientific basis comes from a qualified and respected source (...). It may also rely on the experts to review whether the reasoning (...) is objective and coherent, and whether the particular conclusions drawn by the Member (...) find sufficient support in the evidence. The experts may also be consulted on the relationship between

the risk assessment and the SPS measure in order to assist the panel in determining whether the risk assessment ‘sufficiently warrants’ the SPS measure.⁸³

In the same paragraph, the Appellate Body also warned the panel that it should not seek to determine ‘whether the experts would have done a risk assessment in the same way and *would have reached the same conclusions* as the risk assessor’ (emphasis added).⁸⁴

This brings us to the second issue. The Appellate Body accepted that the investigation into the underlying methodology could be quite intrusive. However, as correctly observed in the literature, if a reviewing body examines the details of underlying methodology, its task is not so different from a body that reviews the substance of evidence. The only difference will be that such a body would concentrate on methodological rather than substantive issues.⁸⁵ The complexity of the required examination (and thus the epistemic inferiority of a panel), as well as the need to rely on specialized expert knowledge, could actually be similar in both cases. A panel may, therefore, end up deciding which methodology is better, which would constitute a situation close to *de novo* review. Consequently an inquiry into methodological aspects does not in itself determine whether we are dealing with a deferential or *de novo* review. If examination is intrusive all concerns raised above with regard to substantive examination of scientific evidence will also apply here.

4. Conclusions

⁸³ Appellate Body, *US – Continued Suspension*, para. 592.

⁸⁴ *Ibidem*.

⁸⁵ Button 2004, p. 186.

The standard of review applicable to complex factual determinations remains ambiguous under the SPS Agreement. The early SPS case law adopted a rather deferential approach that gave a considerable degree of deference to WTO Members. Subsequent jurisprudence opted, however, for a more intrusive examination into the scientific data provided as a justification for national measures. The change in the Appellate Body report in *US/Canada – Continued Suspension* was truly unexpected and had potentially far-reaching implications. Under the new standard, a panel was not expected to determine whether a risk assessment is correct but rather examine its reasonableness. *Australia – Apples* was the first case to elaborate on this new approach. Some of the findings in the Appellate Body report may, however, be disappointing for those who expected to see fairly deferential language. Although the applicable standard of review under the SPS Agreement (and in particular under Articles 2.2 and 5.1) is not *de novo*, the investigation of the WTO panel remains intrusive when assessing the objectivity and coherence of the reasoning included in a contested risk assessment. The same is true with respect to the permissible inquiry into underlying methodology. This may indicate that WTO dispute settlement bodies are unwilling to resign their investigative prerogatives when adjudicating on national SPS measures. This conclusion is not affected by the fact that the standard of review applicable to scientific evidence as such continues to be fairly deferential.

Consequently, one may consider *US/Canada - Continued Suspension* as an anomaly in an otherwise rather consistent line of cases that subscribed to a fairly intrusive standard of review. The *Australia – Apples* report generally followed this trend, however, with one important change relating to the assessment of scientific evidence as such. It was clearly recognized that a panel's task is limited and consists of determining a minimal epistemic status of such evidence. Since the position of the

Appellate Body is quite clear on this issue, one may expect to see the same approach in future panel reports.

Another explanation that could reconcile these two different approaches is that the Appellate Body intends to distinguish between cases on the basis of the risk in question. This would mean that the Appellate Body is willing to apply a more deferential standard of review to trade disputes involving human health issues. *US/Canada - Continued Suspension* is a perfect example of such case with potential risks for human life and health resulting from the consumption of meat from hormone-treated animals. On the other hand, *Australia – Apples* is a traditional phytosanitary case where the values at stake are much lower (namely plant health as opposed to human health). This may justify, in the view of the Appellate Body, a more intrusive examination of a relevant risk assessment. Future reports will most probably shed some more light on the relevance of this distinction. This should allow the more precise identification of future trends in SPS jurisprudence.

References

Bloche, M. Gregg (2002), 'WTO deference to national health policy: Toward an interpretative principle', *Journal of International Economic Law*, 5(4), 825-848.

Bohanes, Jan and Nicolas Lockhart (2009), 'Standard of review in WTO law', in Daniel Bethlehem, Donald McRae, Rodney Neufeld and Issabelle van Damme (eds.), *The Oxford Handbook of International Trade Law*, Oxford: Oxford University Press.

Button, Catherine (2004), *Power to Protect. Trade, Health and World Trade Organization*, Hart Publishing: Oxford and Portland, 378 – 436.

Crawford-Brown, Douglas, Joost Pauwelyn and Kelly Smith (2004), 'Environmental

risk, precaution, and scientific rationality in the context of WTO/NAFTA trade rules’, *Risk Analysis*, 24(2), 461-469.

Christoforou, Theofanis (1999-2000), ‘Settlement of science-based disputes in the WTO: A critical review of the developing case law in the face of scientific uncertainty’, *New York University Environmental Law Journal*, 8(3), 622-648.

Ehlermann Claus-Dieter and Nicolas Lockhart (2004), ‘Standard of review in WTO law’, *Journal of International Economic Law*, 7(3), 491-521.

Epps, Tracey (2011), ‘Recent Developments in WTO Jurisprudence: Has the Appellate Body resolved the issue of an appropriate standard of review in SPS cases?’, unpublished manuscript, on file with the author.

Goh, Gavin (2006), ‘Tipping the apple cart: the limits of science and law in the SPS Agreement after Japan – Apples’, *Journal of World Trade*, 40(2), 655-686.

Gruszczynski, Lukasz (2010), *Regulating Health and Environmental Risks under WTO Law: A Critical Analysis of the SPS Agreement*, Oxford: Oxford University Press.

Guzman, Adrew (2007), ‘Dispute resolution in SPS Cases’, in Dan Horowitz, Daniel Moulis and Debora Steger (eds.), *Ten Years of WTO Dispute Settlement*, London: International Bar Association, pp. 215-233.

Guzman Andrew (2008), ‘Determining the appropriate standard of review in WTO disputes’, paper available at: <http://ssrn.com/abstract=1270894>

Oesch, Matthias (2003), *Standards of Review in WTO Dispute Resolution*, Oxford: Oxford University Press.

Peel, Jacqueline (2010), *Science and Risk Regulation in International Law*, Cambridge: Cambridge University Press.

Prevost, Denise (2005), ‘What role for the precautionary principle in WTO law after

Japan – Apples?’, *EcoLomic Journal of Trade and Environment Studies*, 2(4), 1-14.

Strauss, Peter L., Todd D. Rakoff and Cynthia R. Farina (2003), *Administrative Law: Cases and Comments*, Westbury: Foundation Press.

Thomas, Ryan D., (1999), ‘Where is the beef? Mad cows and the blight of the SPS Agreement’, *Vanderbilt Journal of Transnational Law*, 32(2), 487-517.

Trebilcock, Michael and Julie Soloway (2002), ‘International trade policy and domestic food safety regulation: the case for substantial deference by the WTO Dispute settlement Body under the SPS Agreement’, in Daniel Kennedy and James Southwick (eds.), *The Political Economy of International Trade Law*, Cambridge: Cambridge University Press, pp. 537-574.

Winickoff, David, Sheila Jasanoff, Lawrence Busch, Robin Grove-White and Brian Wynne (2005), ‘Adjudicating the GM food wars: science, risk, and democracy in world trade law’, *Yale Journal of International Law*, 30, 81-123.