

OF APPLES AND ORANGES (AND HORMONES IN BEEF):
SCIENCE AND THE STANDARD OF REVIEW IN WTO
DISPUTES UNDER THE SPS AGREEMENT

JACQUELINE PEEL*

Abstract The standard of review applied by WTO decision-makers in disputes under the Sanitary and Phytosanitary (SPS) Measures Agreement plays a critical role in determining the scope of SPS risk regulatory authority afforded WTO Members by governing the degree to which such measures must be science-based. The standard of review question in SPS jurisprudence finds resonances in international environmental legal adjudication of highly technical disputes, as well as in comparative law concerning judicial review of science-based risk regulation in the United States and European Union. This article considers recent case law of the WTO Appellate Body in *Continued Suspension* and *Australia–Apples* and the extent to which these decisions permit a more deferential approach on the part of WTO decision-makers in evaluating the scientific underpinnings of Members' disputed SPS measures. It is argued that the case law has failed to articulate a strong normative rationale for the current interpretation of the standard of review applied in SPS disputes. Drawing on social scientific findings regarding the limitations of science-based risk assessment in diverse risk settings, the article contends that a more coherent and principled approach to application of the standard of review would allow for its adjustment according to the nature of the risk situation under consideration.

I. INTRODUCTION

The Sanitary and Phytosanitary (SPS) Agreement has attracted academic interest and political controversy since its conclusion as part of the World Trade Organization's (WTO) Uruguay Round agreements in 1995.¹ The Agreement contains novel, science-based requirements that function as the principal mechanism for distinguishing genuine health, environmental and quarantine risk regulatory measures from other trade measures, presumed to be

* Associate Professor, Melbourne Law School, University of Melbourne, Australia. Email: j.peel@unimelb.edu.au. This paper is based on a presentation given in a symposium session on 'The World Trade Organization as a Global Risk Regulator' at the 20th SRA-Europe Meeting, Stuttgart, 8 June 2011. I am grateful for the feedback received from other participants in the symposium including Alberto Alemanno, Lukasz Gruszczynski, Alexia Herwig and Jonathon Wiener. I also benefitted from the comments and feedback of an anonymous peer reviewer of the article.

¹ *Agreement on the Application of Sanitary and Phytosanitary Measures*, opened for signature 15 April 1994, 1867 UNTS 493 (in force 1 January 1995) (SPS Agreement).

motivated by protectionism.² If a dispute arises over a Member's measures, it may be taken before the WTO dispute settlement body where legal decision-makers (panels and a standing Appellate Body) evaluate questions such as the 'sufficiency' of the scientific evidence underlying a measure and its relationship to a science-based risk assessment. Potentially these provisions may require international judicial bodies to undertake a searching review of the scientific underpinnings of SPS risk regulation, with the potential to place significant constraints on the scope of domestic regulatory autonomy in this field.

While many found the first SPS dispute brought under the Agreement—that of *Hormones* in 1998—reassuring given the efforts made by the Appellate Body to articulate a flexible notion of risk assessment applicable under the SPS Agreement,³ later decisions evolved in a direction that gave a large role to science (and experts advising panels)⁴ in evaluating the international legal legitimacy of SPS measures. In November 2008, the WTO Appellate Body issued its decision in a further round of the *Hormones* dispute (this time examining whether amended European measures complied with the earlier Appellate Body rulings, warranting the suspension of trade sanctions).⁵ The *Continued Suspension* decision articulated new requirements around the so-called 'standard of review' to be applied by panels in evaluating the scientific basis and adequacy of a Member's risk assessment. The standard of review governs the extent of the investigative authority enjoyed by WTO dispute settlement bodies when it comes to examining disputed facts, including scientific evidence. In the wake of the *Continued Suspension* case, some commentators saw this decision as an indication in WTO dispute settlement of (a return to) greater deference to the scientific judgment of domestic authorities in SPS risk regulation.⁶

Two years later in November 2010 the Appellate Body issued its first ruling since *Continued Suspension* in the case of *Australia–Apples*.⁷ The Australian appeal squarely raised the question of whether the panel applied the correct standard of review in evaluating the relevant risk assessment. The Appellate Body's interpretation of its findings in *Continued Suspension* suggests that

² Joost Pauwelyn, 'The WTO Agreement on Sanitary and Phytosanitary (SPS) Measures as Applied in the First Three SPS Disputes EC – Hormones, Australia – Salmon and Japan Varietals' (1999) 2 JIEL 644.

³ *European Communities – Measures Concerning Meat and Meat Products*, Report of the Appellate Body, WT/DS26/AB/R & WT/DS48/AB/R (16 January 1998).

⁴ Article 11.2 of the SPS Agreement provides: 'In a dispute under this Agreement involving scientific or technical issues, a panel should seek advice from experts chosen by the panel in consultation with the parties to the dispute.'

⁵ *United States – Continued Suspension of Obligations in the EC – Hormones Dispute*, Report of the Appellate Body WT/DS320/AB/R, 16 October 2008 (*Continued Suspension*).

⁶ Lukasz Gruszczynski, *Regulating Health and Environmental Risks Under WTO Law: A Critical Analysis of the SPS Agreement* (2010) 51–2; Sungjoon Cho, 'United States – Continued Suspension of Obligations in the EC – Hormones Dispute' (2009) 103 AJIL 302.

⁷ *Australia – Measures Affecting the Importation of Apples from New Zealand*, WT/DS367/AB/R, Report of the Appellate Body, 29 November 2010 (*Australia–Apples*).

despite the articulation of a new standard, its application may still result in detailed scrutiny of Members' risk assessment findings, at least in the context of SPS disputes raising quarantine risk issues.

What the *Australia–Apples* rulings bring to the fore are problems with treating all SPS risk situations alike as regards the stringency of the review standard applied. The SPS Agreement covers a diverse range of human health, food safety and quarantine risks,⁸ and potentially also extends to some environmental risks.⁹ As all SPS risk situations are not alike, this article argues there is a need to differentiate between the stringency of the standard of review applicable in different risk situations. Frameworks developed in the social scientific literature for understanding the limitations of conventional science-based risk assessment can be adapted to supply key parameters for differentiating risk situations on the basis of the extent of scientific uncertainty and the level of socio-political contestation with respect to risks.¹⁰ This approach could provide a principled and more transparent method for the use of science in evaluating WTO Members' risk assessments and discerning legitimate SPS measures. In turn, practice in the WTO SPS context may inform broader questions regarding the use of science and expert evidence in international dispute settlement involving environmental risks, questions which are receiving increased attention.¹¹

Part II of the article highlights the multifaceted role of the standard of review in SPS disputes and discusses how questions over the appropriate stringency of review have been approached, both in the setting of the SPS Agreement and its jurisprudence, and in comparative contexts such as the judicial review of risk regulation in the United States (US) and European Union (EU). Part III discusses the deficiencies of a stringent science-based review of risk regulatory measures in dealing with the diversity of possible risk situations covered by the SPS Agreement and draws on established frameworks in the social scientific literature to articulate other important parameters for application of the standard of review. Parts IV and V then examine the clarifications of the standard of review offered in the WTO Appellate Body's *Continued Suspension* and *Australia–Apples* decisions, analysing whether these cases allow scope for a standard of review differentiated according to the risk situation under consideration. Finally, Part VI of the article turns to the question of how a differentiated standard of review, looking to parameters of scientific uncertainty and socio-political contestation regarding risks, might operate in the

⁸ SPS Agreement, Annex A.1 defines the SPS measures covered by the Agreement.

⁹ Jacqueline Peel, 'A GMO by Any Other Name ... Might Be an SPS Risk!: Implications of Expanding the Scope of the WTO Sanitary and Phytosanitary Measures Agreement' (2006) 17 EJIL 1009.

¹⁰ These frameworks and the parameters of scientific uncertainty and socio-political contestation are discussed further in part III below.

¹¹ Philippe Sands, 'Water and International Law: Science and Evidence in International Litigation' (2010) 22 ELM 15.

practice of WTO SPS dispute settlement, with potentially broader significance for cases concerning environmental risk measures arising in other areas of international law.

II. SCIENCE AND THE STANDARD OF REVIEW

A. Role of the Standard of Review

The standard of review applied by WTO panels in disputes under the SPS Agreement—while seemingly a technical question—is in practice intimately tied up with the extent to which the SPS Agreement limits the risk regulatory autonomy of WTO Members in this field.¹² The SPS Agreement contains the concept of a Member's 'appropriate level of protection' (ALOP), otherwise referred to as the notion of 'acceptable risk', which is defined as the level of protection deemed appropriate by a WTO Member establishing an SPS measure to protect human, animal, or plant life or health within its territory.¹³ The WTO Appellate Body has, on a number of occasions, stressed that it is the prerogative of Members (and not the WTO) to set ALOPs,¹⁴ an approach consistent with conventional notions of risk regulation that designate decisions about the levels of risk a society is prepared to accept as a socio-political (and not a scientific) matter.¹⁵

Nonetheless, in adopting SPS measures designed to ensure an ALOP is achieved, WTO Members must comply with the scientific and risk assessment provisions of the SPS Agreement. In particular, a WTO Member—if not basing its measures on international standards,¹⁶ or where no such international standards exist—shall ensure that its SPS measure is 'applied only to the extent necessary to protect human, animal or plant life or health, is based on scientific principles and is not maintained without sufficient scientific evidence'.¹⁷ It must also ensure that the SPS measure is 'based on' a risk assessment 'as appropriate to the circumstances'¹⁸ that takes into account specified factors, including scientific evidence.¹⁹

¹² Catherine Button, *The Power to Protect: Trade, Health and Uncertainty in the WTO* (Hart Publishing, Oxford, 2004), 193.

¹³ SPS Agreement, Annex A 5 and accompanying note.

¹⁴ Michael Ming Du, 'Autonomy in Setting Appropriate Level of Protection under the WTO Law: Rhetoric or Reality?' (2010) 13 *JIEL* 1077. However, the right of a WTO Member to set its own ALOP is not entirely without limits as is evident from Articles 5.4–5.6 of the SPS Agreement.

¹⁵ Royal Society Study Group, *Risk: Analysis, Perception and Management* (Royal Society, London, 1992).

¹⁶ SPS Agreement, Articles 3.1 and 3.3. International standards are those promulgated for a particular risk issue by the Codex Alimentarius (dealing with food safety), the International Office of Epizootics (dealing with animal health) and the International Plant Protection Convention (dealing with plant health). Domestic measures which conform to relevant international standards are deemed to protect human, animal or plant life or health and are presumed to be consistent with the requirements of the SPS Agreement, as well as those of the General Agreement on Tariffs and Trade (GATT): Article 3.2.

¹⁷ SPS Agreement, Article 2.2.

¹⁸ SPS Agreement, Article 5.1.

¹⁹ SPS Agreement, Articles 5.2 and 5.3.

In past case law the Appellate Body has indicated that there is a close relationship between the scientific evidence and risk assessment requirements of the SPS Agreement, and that the relevant provisions must be ‘constantly’ read together.²⁰ Moreover, scientific evidence will only be considered insufficient (which under the SPS Agreement is a justification for the adoption of provisional, precautionary measures)²¹ when a WTO Member is unable to perform ‘an adequate assessment of risks’ as required by the Agreement on the basis of that evidence.²²

In this context, the standard of review applied in WTO dispute settlement in evaluating a contested SPS measure plays a critical function in mediating between the prerogative of WTO Members to set their own acceptable risk levels and compliance with the requirements set down for legitimate measures in the SPS Agreement. A very lenient application of the standard of review would allow WTO Members significant regulatory autonomy in establishing SPS risk measures but could also forgo the expected benefits of the SPS Agreement that come from tying the adoption of such measures to certain standards of ‘rational’ decision-making.²³ On the other hand, a very intrusive application of the standard of review might constitute WTO decision-makers not only as the judge of the adequacy of the underlying science but also place significant constraints on what kind of SPS measures can be adopted by Members, undermining their choice of an acceptable level of risk.

1. Role of experts versus ‘other’ risk perspectives in SPS review

The stringency of the standard of review adopted by WTO decision-makers in SPS disputes is also highly influential in determining the types of information and perspectives on risk that will be considered (or privileged) in SPS review. For instance, the standard of review adopted is an important factor in determining the extent of the role played by scientific experts in SPS dispute settlement.²⁴ Given the technical nature of many SPS disputes, and the lack of scientific expertise possessed by panels, the appointment of experts to advise panels has been a constant feature of the SPS jurisprudence.²⁵ To the extent that panels adopt a more intrusive standard of review in examining the facts of

²⁰ *Hormones*, para 180. See also *Continued Suspension*, para 674.

²¹ SPS Agreement, Article 5.7. The Appellate Body has stated that the general international environmental law principle of precaution finds reflection in this provision: *Hormones*, para 124; see also *Continued Suspension*, para 680.

²² *Japan – Measures Affecting the Importation of Apples*, Report of the WTO Appellate Body, WT/DS245/AB/R, 26 November 2003 (*Japan–Apples*), para 179.

²³ Warren H Maruyama, ‘A New Pillar of the WTO: Sound Science’ (1998) 32 *IntLAW* 651.

²⁴ Lukasz Gruszczynski, ‘The Use of Experts in Environmental and Health Related Trade Disputes: Processes and Outcomes’, forthcoming in an edited book on experts and legitimacy to be published by Cambridge University Press in 2012 (paper on file with author).

²⁵ Joost Pauwelyn, ‘The Use of Experts in WTO Dispute Settlement’ (2002) 51 *ICLQ* 325.

a dispute, their dependence on advising experts is likely to increase in order to permit them to understand and evaluate the scientific arguments put forward by WTO Members. In turn this will tend to amplify the importance of the panel's decisions about how experts are selected, consulted and the nature of the questions put to them, as well as issues around the diversity of experts consulted in a dispute, their level of independence and impartiality, and the use made of their expertise in panel decision-making.

In the area of international environmental dispute settlement more broadly, similar questions are attracting increasing interest given the equally technical nature of many environmental disputes and hence the difficulties international judicial tribunals face in evaluating complex factual information.²⁶ In the *Pulp Mills* decision of the International Court of Justice, the Court commented on the treatment to be afforded to expert evidence and the methodology to be applied by the Court in considering the 'vast amount' of scientific and technical information before it.²⁷ A majority of the Court was of the view that 'despite the volume and complexity of the factual information submitted to it, it is the responsibility of the Court, after having given careful consideration to all the evidence placed before it by the Parties, to determine which facts must be considered relevant, to assess their probative value, and to draw conclusions from them as appropriate';²⁸ an approach that bears many similarities to the SPS 'objective assessment' standard. However, dissenting Judges Al-Khasawneh and Simma criticized the majority for clinging 'to the habits it has traditionally followed for the assessment and evaluation of evidence', which might only serve to 'increase doubts in the international legal community whether [the Court], as an institution, is well-placed to tackle complex scientific questions'.²⁹ They argued that the Court should have availed itself of powers to have recourse to outside sources of expertise in handling the complex scientific and technical dispute before it.³⁰ Significantly, they pointed to the WTO dispute settlement system practice of consulting and questioning experts in SPS disputes as an example, in their view, of international 'best practice'.

²⁶ See particularly, Caroline E Foster, *Science and the Precautionary Principle in International Courts and Tribunals: Expert Evidence, Burden of Proof and Finality* (CUP, Cambridge, 2011). See also Shabtai Rosenne, 'Fact-Finding Before the International Court of Justice', in *Essays on International Law and Practice* (Brill, Leiden, 2007) 235; Anna Riddell and Brendan Plant, *Evidence before the International Court of Justice* (BIICL, London, 2009) 353; Clive M Schofield and Chris H Carleton, 'Technical Considerations in Law of the Sea Dispute Resolution', in Alex G Oude Elferink and Don R Rothwell (eds), *Oceans Management in the 21st Century* (Brill, Leiden, 2004) 239ff.

²⁷ *Pulp Mills on the River Uruguay (Argentina v. Uruguay)* (Judgment) 20 General List No 135, 20 April 2010 (*Pulp Mills*) para 165. The majority expressed some qualms about the reliability of evidence given by experts appearing as counsel for one or other of the parties and a preference for independent expert testimony: para 167.

²⁸ *Pulp Mills*, majority judgment, para 168.

²⁹ *Pulp Mills*, Joint Dissenting Opinion, Judges Al-Khasawneh and Simma, para 3.

³⁰ *ibid* paras 7–10.

The importance of robust procedures relating to the presentation of expert evidence is all the more important in the light of social scientific research indicating that experts bring to the exercise of evaluation their own implicit understandings and ‘framings’ of the SPS or environmental risks at issue,³¹ which may not always accord with the risk framings adopted by the WTO Member or state whose measures are in dispute.³² An intrusive panel review of a WTO Member’s risk assessment, drawing heavily on expert views (especially if these views are accepted uncritically), thus risks supplanting the risk framings considered important by advising experts for those deemed important by the WTO Member in carrying out the risk assessment.

A similar point can be made in relation to the scope for the inclusion of other, non-scientific, viewpoints in a WTO Member’s risk assessment exercise, such as the extent of public concern over a particular risk. The WTO Appellate Body in *Hormones* indicated that SPS risk assessment is not confined to an evaluation of ‘risk ascertainable in a science laboratory operating under strictly controlled conditions’ but also extends to ‘risk in human societies as they actually exist, in other words, the actual potential for adverse effects on human health in the real world where people live and work and die’.³³ This broader approach to risk assessment appears to take account of perspectives on risk offered by social scientists in the ‘constructivist’ school, who see value judgments and social processes as inherent to the process of risk assessment.³⁴ It also accords more closely with notions of environmental impact assessment in international law that generally allow for a range of information inputs beyond scientific views, including public comment.³⁵

One way to take account of non-scientific inputs, including public opinion, in the review of SPS measures is via adjusting the standard of review to place less emphasis on the need for a strong correlation between the measures and the underlying scientific evidence.³⁶ In this sense, the standard of review can become a key mechanism under the SPS Agreement allowing for a diversity of approaches to understanding and evaluating SPS risks. Conversely, application of a more stringent standard of review, which takes a close look at the scientific

³¹ For an overview of the pertinent findings of this research see Sheila Jasanoff, ‘What Judges Should Know About the Sociology of Science’ (1993) 77 *Judicature* 2, 77. Decisions about how to ‘frame’ a risk assessment ie what hazards to include, what evidence to consider and so on are critical in determining the scope of the risk assessment conducted.

³² In *Continued Suspension*, para 686, the Appellate Body recognized that a Member’s ALOP plays a role in the ‘scope or method’ of the risk assessment undertaken to justify SPS measures.

³³ *Hormones*, para 187. See also *Continued Suspension*, para 527.

³⁴ Jacqueline Peel, *Science and Risk Regulation in International Law* (CUP, Cambridge, 2010) 104.

³⁵ Neil Craik, *The International Law of Environmental Impact Assessment* (CUP, Cambridge, 2008) 196–8.

³⁶ Nonetheless, inclusion of public opinion in SPS risk assessment remains highly controversial: Joanne Scott, ‘European Regulation of GMOs: Thinking About “Judicial Review” in the WTO’ in Michelle Everson and Ellen Vos (eds), *Uncertain Risks Regulated* (Routledge-Cavendish, Milton Park, 2009) 317.

justification for a WTO Member's measures, is likely to foreclose possibilities for the acceptance of non-scientific inputs and public opinion as elements relevant to the review of risk assessment under WTO SPS rules.

2. Interpretation of the standard of review

The text of the SPS Agreement provides no ready solution to the question of the appropriate stringency of review to be applied by panels and the WTO Appellate Body in SPS disputes. The term 'standard of review' is not found in the SPS Agreement and no other specific method of evaluation of evidence is required for disputes under the Agreement.³⁷ In *Hormones*, the European Communities sought to argue that the appropriate standard was one of 'deferential reasonableness', limiting the panel to assessing the propriety and objectivity of the Communities' risk assessment process. The Appellate Body, however, indicated that in understanding the standard of review applicable in SPS cases the correct starting point was Article 11 of the Dispute Settlement Understanding (DSU), applicable across all domains of WTO dispute settlement.³⁸ The key part of Article 11 provides that in a dispute under one of the WTO Agreements 'a panel should make an objective assessment of the matter before it, including an objective assessment of the facts of the case and the applicability of and conformity with the relevant covered agreements'.

The only other clarifications of the 'objective assessment' standard offered by the WTO Appellate Body in *Hormones* went to the question of what is *not* permitted under this standard of review.³⁹ The Appellate Body held that the standard is neither that of *de novo* review nor total deference.⁴⁰ By *de novo* review, the Appellate Body contemplated a situation where the panel reaches its own assessment of the risks, based on the scientific material put forward by the parties and the panel's advising experts.⁴¹ On the other hand, 'total deference' connoted a situation where the panel simply adopts the same evaluation of the scientific evidence as the Member concerned.⁴² The WTO Appellate Body's judgment indicated that the applicable standard of review falls somewhere between these two extremes but did not offer any further specification.

³⁷ This differs from other WTO Agreements, such as the Anti-Dumping Agreement, Article 17.6(i), which articulates a specific standard of review for disputes under that Agreement based on panels determining 'whether the authorities' establishment of the facts was proper and whether the evaluation of those facts was unbiased and objective'.

³⁸ *Marrakesh Agreement Establishing the World Trade Organization*, opened for signature 15 April 1994, 1867 UNTS 3 (in force 1 January 1995), annex 2 (*Understanding on Rules and Procedures Governing the Settlement of Disputes*) 1869 UNTS 401 (DSU).

³⁹ Claus-Dieter Ehlermann and Nicolas Lockhart, 'Standard of Review in WTO Law' (2004) 7 JIEL 491.

⁴¹ The Appellate Body described *de novo* review as 'complete freedom to come to a different view than the competent authority of the Member whose act or determination is being reviewed': *ibid.*, para 111.

⁴² See *Japan-Apples*, para 165.

B. Comparative Experience with Judicial Review of Risk Regulation

As demonstrated by the experience of judicial review of risk regulation from other comparative jurisdictions, such as jurisprudence at the federal level in the US and at the supranational level in the EU, locating the ideal stringency of review of science-based risk regulation on the spectrum between *de novo* review and total deference is not an easy task.

In the US, where risk regulations introduced by federal executive agencies have been challenged as insufficiently science-based by regulated industries, the review standard employed has been based, since the 1970s, on the 'hard look' doctrine.⁴³ In recognition of the difficulties presented by the review of technical decision-making, this doctrine professes a policy of judicial restraint with respect to evaluation of the science underlying a regulatory measure, but with a 'hard look' at the policy choices made on the basis of that evidence.⁴⁴

In practice, the US federal case law reveals a variety of understandings of how to apply the hard look standard in judicial review cases. This divergence was exemplified in the late 1970s in the judicial writings of two justices of the Court of Appeals for the DC Circuit. Judge Leventhal was a prominent advocate of reviewing judges scrutinizing the substantive underpinnings of an executive agency's decision to determine whether its exercise of discretion was reasonable.⁴⁵ By taking a 'hard look' at the agency's record and reasoning supporting a decision, courts following this approach often undertook a searching review of the underlying scientific evidence. In contrast, Chief Judge Bazelon believed that reviewing judges should 'scrutinize agency proceedings with extreme care' focusing on ensuring that agency procedures were adequate to allow for public participation in rule-making and full disclosure of areas of scientific uncertainty.⁴⁶ Over time, court rulings following Chief Judge Bazelon's lead devised increasingly more stringent procedural requirements to be met by agencies in order to satisfy the demands of the 'hard look' doctrine,⁴⁷ although, ultimately, this did not guarantee judicial restraint since judges often needed to probe the science underlying measures to determine whether additional procedures were necessary.⁴⁸ Further, procedural obligations imposed by reviewing courts on agency risk regulation drew criticism for unnecessarily complicating the federal regulatory process without improving transparency in agencies' use of science,⁴⁹ and were eventually overruled

⁴³ Joel A Tickner and Sara Wright, 'The Precautionary Principle and Democratizing Expertise: a US Perspective' (2003) 30 *Science and Public Policy* 213.

⁴⁴ I am grateful to Jonathon Wiener for this succinct articulation of the doctrine.

⁴⁵ *Ethyl Corporation v EPA*, 541 F 2d 1, 68-9 (1976).

⁴⁶ David L Bazelon, 'Science and Uncertainty: A Jurist's View' (1981) 5 *HarvEnvtlLRev* 212.

⁴⁷ See, eg *Natural Resources Defense Council v Nuclear Regulatory Commission*, 547 F 2d 633 (1976).

⁴⁸ Sheila Jasanoff, *Science at the Bar* (Harvard University Press, Cambridge, MA, 1995), 77.

⁴⁹ Wendy E Wagner, 'The Science Charade in Toxic Risk Regulation' (1995) 95 *ColumLRev* 1613.

by the US Supreme Court, albeit while leaving the ‘hard look’ doctrine intact.⁵⁰ Federal courts thus remained free to scrutinize the scientific underpinnings of agencies’ risk regulatory measures, leading to the development of increasingly stringent requirements for agency risk assessment.⁵¹

In the EU, the issue of the stringency of the review to be applied by European courts examining risk regulatory measures has arisen principally in cases challenging measures adopted by the Union institutions on the basis of the precautionary principle.⁵² In general, in cases involving review of the competence of Union institutions to adopt risk regulations in complex and technical cases, the standard of review applied by European courts has been deferential.⁵³ However, the European jurisprudence—like that in the US—has demonstrated an increasing tendency to tighten requirements around the adoption of precautionary risk regulations. This tightening has largely taken a procedural direction,⁵⁴ in the sense of specifying rigorous standards around the pre-requisites for decision-making, the quality of the scientific advice supplied to the Union institutions to support decision-making, and the need for transparency and reason-giving.⁵⁵

Even so, European Courts have not been prepared to confirm the absolute discretion of Union institutions to regulate health and environmental risks in

⁵⁰ *Vermont Yankee Nuclear Power Corp. v Natural Resources Defense Council*, 435 U.S. 519 (1978).

⁵¹ See, eg the Supreme Court decision in *Industrial Union Department, AFL-CIO v American Petroleum Institute*, 448 U.S. 607 (1980). See also Thomas McGarity, ‘The Courts and the Ossification of Rulemaking: A Response to Professor Seidenfeld’ (1997) 75 *TexLR* 525.

⁵² Alberto Alemanno, ‘Science and EU Risk Regulation: The Role of Experts in Decision-Making and Judicial Review’, in Ellen Vos (ed), *European Risk Governance: its Science, its Inclusiveness and its Effectiveness*, CONNEX Report Series No 6, 37. In EU law the precautionary principle is construed as follows: ‘where there is uncertainty as to the existence or extent of risks to human health, protective measures may be taken without having to wait until the reality and seriousness of those risks become fully apparent’: Case C-236/01, *Monsanto Agricoltura Italia* [2003] ECR I-8105, para 111. See also Case C-180/96, *United Kingdom v Commission* [1998] ECR I-2265, para 99; Case C-157/96 *National Farmers’ Union and Others* [1998] ECR I-2211, para 63; Case E-3/00 *EFTA Surveillance Authority v Norway* [2001] EFTA Ct Rep 73, para 31.

⁵³ The judiciary ‘must confine itself to ascertaining whether the exercise by the institutions of their discretion in that regard is vitiated by a manifest error or a misuse of powers or whether the institutions clearly exceeded the bounds of their discretion’: Case T-13/99, *Pfizer Animal Health SA v Council of the European Union* [2002] ECR II-3305 (*Pfizer*), para 169. Judicial deference to discretion in decision-making in the context of scientific uncertainty has a long tradition in EU law. See, eg Case C-331/88, *R v Minister of Agriculture, Fisheries and Food and Secretary of State for Health ex parte Fedesa* [1990] ECR I-4023.

⁵⁴ Joanne Scott, ‘International Trade and Environmental Governance: Relating Rules (and Standards) in the EU and the WTO’ (2004) 15 *EJIL* 307.

⁵⁵ See particularly, *Pfizer*, (n 53); Case T-70/99 *Alpharma Inc. v Council of the European Union* [2002] ECR II-3495 (*Alpharma*); and Case C-79/09, *Gowan Comércio Internacional e Serviços Lda v. Ministero della Salute*, Judgment of the Court of Justice (Second Chamber) of 22 December 2010. For analysis of the broadly deferential approach taken by the European Court of Justice in the latter case see Alberto Alemanno, ‘Case C-79/09, *Gowan Comércio Internacional e Serviços Lda v. Ministero della Salute*, Judgment of the Court of Justice (Second Chamber) of 22 December 2010’ (2011) 48 *CML Rev* 1329.

conditions of scientific uncertainty, perhaps out of a concern that these institutions lack the necessary credibility and social legitimacy to carry off this level of discretion.⁵⁶ In these circumstances, we might expect to see a trend over time in European risk regulation towards a greater degree of scientific rigour and increased reliance on risk assessments.⁵⁷

The body of comparative risk regulatory jurisprudence that has emerged in the US and EU illustrates intermediate points along the standard of review spectrum between *de novo* review and total deference. The 'hard look' doctrine, as it has evolved in the US might be placed more towards the *de novo* end of this spectrum, whereas the approach of the European courts has been more deferential. At the same time this experience illustrates the difficulties courts face in seeking to ensure a rational, scientifically sound basis for decision-making while still allowing authorities sufficient autonomy to introduce risk regulations to respond to uncertain or socially unacceptable risks. These difficulties are all the more acute for WTO judicial bodies given the lesser degree of legitimacy they enjoy compared to their national, and even supranational EU-level, counterparts.⁵⁸

C. A Trend towards Quasi De Novo Review in the SPS Jurisprudence

The lack of clarity around the standard of review applicable in WTO SPS disputes contributed to a trend in cases following the *Hormones* decision towards an approach coming close to *de novo* review. The foundations for that trend were laid in the *Hormones* case itself. While in *Hormones* the Appellate Body indicated some flexibility around the use of scientific evidence in a Member's risk assessment (for instance, stating that the science relied upon did not have to constitute the majority view of the relevant scientific community)⁵⁹ it also ruled that SPS measures must have a 'rational relationship' to a science-based risk assessment,⁶⁰ and that in conducting a risk assessment a WTO Member must be able to produce 'specific' scientific studies substantiating the risk of concern.⁶¹ These two requirements, and in particular the latter, encourage panels to look closely at the scientific evidence to determine

⁵⁶ Veerle Heyvaert, 'Facing the Consequences of the Precautionary Principle in European Community Law' (2006) 31 EL Rev 203.

⁵⁷ Some commentators already see evidence of a convergence in US and EU risk regulatory approaches with supranational/WTO review as an important driver: David Vogel, 'The Hare and the Tortoise Revisited: The New Politics of Consumer and Environmental Regulation in Europe' (2003) 33 British J Pol Sci 567.

⁵⁸ Jacqueline Peel, 'Risk Regulation under the WTO SPS Agreement: Science as an International Normative Standard?' (2004) Jean Monnet Working Paper No 2/2004.

⁵⁹ *Hormones*, para 194. See also *Continued Suspension*, para 529.

⁶⁰ *Hormones*, para 193.

⁶¹ *Hormones*, para 200. See also *Continued Suspension*, para 530. The Appellate Body in *Continued Suspension* indicated some loosening of this requirement, particularly in the case of cumulative risks: see paras 562–3.

whether the studies put forward by a defending Member indicate a sufficient basis for a finding of risk and the implementation of risk regulatory measures. The danger of such an approach is that it can lead to scientific risk findings being ‘assigned relevance to the exclusion of everything else’.⁶²

The Appellate Body did indicate, however, that a less stringent standard of risk assessment might apply in the context of some SPS risks, such as food safety risks, than in relation to other SPS risks, such as those of quarantine significance arising from the potential introduction, establishment and spread of pests and diseases in a WTO Member’s territory. The basis for this distinction was the text of the definition of ‘risk assessment’ in the SPS Agreement, which speaks of an ‘evaluation of the *potential* for adverse effects on human or animal health’ in the case of food-borne risks and an ‘evaluation of the *likelihood* of entry, establishment or spread of a pest or disease’ in the case of quarantine risks.⁶³ In *Hormones*, the Appellate Body interpreted ‘potential’ for adverse effects to mean simply the possibility of harm;⁶⁴ whereas ‘likelihood’ was equated in the later case of *Salmon* with a requirement for the assessment of probabilities.⁶⁵ As discussed further in Part V, whatever the textual reasoning, the normative rationale for this distinction is not compelling and can lead to some odd results in practice.

The cases that followed *Hormones* in the ensuing 15 years—*Salmon* (1998), *Varietals* (1999), *Japan–Apples* (2003), *Biotech* (2006) and the panel decisions in *Continued Suspension* (2008) concerned both quarantine/environmental risks and food safety risks, yet stringency scrutiny of WTO Member’s risk assessments and the underlying science was a constant feature. It is not possible here to give a comprehensive account of the development of the SPS jurisprudence.⁶⁶ For present purposes, however, a number of findings in these cases are pertinent in illustrating an overall trend in the SPS jurisprudence towards a more stringent review of the scientific underpinnings of risk regulatory measures. These findings include the following:

- Quarantine risk assessment concerning the likelihood of pest or disease entry, establishment and spread in a WTO Member’s territory requires an evaluation of risk ‘probabilities’ and not mere ‘possibilities’ (*Salmon*). Although in risk-assessment practice probabilities of risk are usually evaluated in *quantitative* terms (ie a one in a million risk of x), the Appellate Body has held that a *qualitative* assessment is also acceptable (ie evaluating risk as low, medium, high etc).⁶⁷

⁶² As the Appellate Body itself acknowledged in *Hormones*, para 193.

⁶³ SPS Agreement, Annex A 4.

⁶⁴ *Hormones*, para 184.

⁶⁵ *Australia – Measures Affecting Importation of Salmon*, Report of the WTO Appellate Body, WT/DS18/AB/R, 20 October 1998 (*Salmon*), para 173.

⁶⁶ For further details see Peel, (n 34) ch 5.

⁶⁷ *Salmon*, paras 124 and 128.

- The requirement of Article 2.2 of the SPS Agreement for ‘sufficient’ scientific evidentiary support for a Member’s SPS measure necessitates a ‘rational relationship’ between the measure and the underlying scientific evidence (*Varietals*),⁶⁸ emphasizing a panel’s assessment of the scientific underpinnings of SPS measures in evaluating their legal legitimacy.⁶⁹
- A Member’s measures may be ruled ‘disproportionate’ and illegal under the SPS Agreement where experts advising the panel consider the available scientific evidence establishes the existence of ‘negligible risk’ (*Japan–Apples*).⁷⁰
- A risk assessment of the SPS matter at issue, conducted at another level of government, which finds no unacceptable risks can be used to call into question the validity of scientific evidence reaching an opposite conclusion (*Biotech*),⁷¹ notwithstanding that the assessments may have adopted different framings of the risks.⁷²

The trend in the SPS jurisprudence toward heavy reliance on scientific information (largely provided by the experts advising the panel) came full circle with the panel reports in the *Continued Suspension* case.⁷³ The panel was called upon to consider whether new scientific studies on the health effects of hormone residues in meat products—undertaken following the EU’s loss in the first *Hormones* dispute—amounted to a risk assessment sufficient to warrant the EU’s amended measures that maintained bans on hormone-treated beef. The methodology of the panel in approaching this task was essentially one of surveying and summarizing the opinions of its expert advisors on various questions. The majority view to emerge from this process was assumed to reflect the ‘best science’ and the basis upon which the EU measures should be judged.⁷⁴

⁶⁸ *Japan – Measures Affecting Agricultural Products*, Report of the WTO Appellate Body, WT/DS76/AB/R, 22 February 1999 (*Varietals*), para 84.

⁶⁹ Oliver Landwehr, ‘Decisions of the Appellate Body of the World Trade Organization: Japan – Measures Affecting Agricultural Products’ (1999) 10 EJIL 803.

⁷⁰ *Japan – Measures Affecting the Importation of Apples*, Report of the Panel, WT/DS245/R, 15 July 2003 (*Japan–Apples* Panel Report), para 8.198. Despite the fact that this ruling calls into question whether it was Japan or the panel determining an acceptable level of risk, on appeal the Appellate Body affirmed the Panel’s methodology noting it did ‘not exhaust the range of methodologies available to determine whether a measure is maintained “without sufficient scientific evidence” within the meaning of Article 2.2’: *Japan–Apples*, para 164.

⁷¹ *European Communities – Measures Affecting the Approval and Marketing of Biotech Products*, Reports of the Panel, WT/DS291/R, WT/DS292/R and WT/DS293/R, 29 September 2006 (*Biotech*), para 8.9.

⁷² Alexia Herwig, ‘Whither Science in WTO Dispute Settlement?’ (2008) 21 LJIL 834.

⁷³ Identical reports were issued in the dispute brought by the US as in that brought by Canada: see *US – Continued Suspension of Obligations in the EC – Hormones Dispute*, Report of the Panel, WT/DS320/R, 31 March 2008 (*US–Continued Suspension* Panel Report), [7.612]; *Canada – Continued Suspension of Obligations in the EC – Hormones dispute*, Report of the Panel, WT/DS321/R, 31 March 2008 (*Canada – Continued Suspension* Panel Report).

⁷⁴ *Continued Suspension*, paras 597, 598 and 602.

III. DIFFERENTIATING RISK SITUATIONS (APPLES FROM ORANGES...)

At this point a reader might well ask why we should be troubled by the emergence of a stringent science-based review of SPS measures given that the SPS Agreement itself places heavy emphasis on the role of science in discerning between legitimate and illegitimate domestic risk regulations. In this context, stringent review of the scientific underpinnings of SPS measures might be regarded as the best means of ensuring an 'objective assessment' of the facts in dispute as required by the WTO DSU. However, this approach assumes a monolithic concept of risk, and indeed science, which employs the latter as a litmus test for the legitimacy of assessments of the former. A one-dimensional approach of this kind does in fact underlie conventional science-based processes of risk assessment. According to this view, risk assessment is a scientific process of identifying potential hazards or adverse outcomes, and evaluating their likelihood of occurrence in order to come up with an aggregated risk characterization.⁷⁵

Over time this view of risk assessment—or rather its capacity to capture all dimensions of risk—has come under increasing challenge, particularly from social scientists working in the constructivist tradition. While conventional notions of science and risk stress objectivity, certainty and universality,⁷⁶ constructivist notions highlight the many uncertainties in scientific research and the importance of social processes in contributing to our understanding of what constitutes scientific knowledge and risk.⁷⁷ Building on these critiques, a number of researchers have sought to develop frameworks for defining the circumstances in which 'normal science'⁷⁸ and conventional risk assessment present useful tools for understanding risk, and those circumstances in which the application of such tools can lead to a distorted view of risk.⁷⁹ The underlying premise of such frameworks is that the use of science in risk

⁷⁵ John Adams, *Risk* (UCL Press, London, 1995), 8.

⁷⁶ Ortwin Renn, 'Concepts of Risk: A Classification' in Sheldon Krinsky and Dominic Golding (eds), *Social Theories of Risk* (Praeger Publishers, Westport, CT, 1992) 53.

⁷⁷ Tom Horlick-Jones and Jonathan Sime, 'Living on the Border: Knowledge, Risk and Transdisciplinarity' (2004) 36 *Futures* 447.

⁷⁸ The term 'normal science' was famously employed by Thomas Kuhn to describe the standard business of scientific research within a given (unchallenged) paradigm: Thomas S Kuhn, *The Structure of Scientific Revolutions* (3rd edn, University of Chicago Press, Chicago, IL, 1996).

⁷⁹ Examples include Andreas Klinke and Ortwin Renn, 'A New Approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based Strategies' (2002) 22 *Risk Analysis* 1071; Ortwin Renn et al, *The Application of the Precautionary Principle in the European Union* (Precaupri Project, European Commission, Stuttgart, 2003); Andrew Stirling, 'On Science and Precaution in the Management of Technological Risk, vol I: A Synthesis Report of Case Studies' (European Science and Technology Observatory, 1999); Andrew Stirling, 'On Science and Precaution in the Management of Technological Risk, Volume II: Case Studies' (European Science and Technology Observatory, 2001).

decision-making should 'focus . . . on the conditions under which it is valid, and whether those conditions prevail in the situation of interest'.⁸⁰

Many of these frameworks have their basis in the seminal work of Silvio Funtowicz and Jerome Ravetz on 'post-normal' science. Post-normal science is a way of doing science said to be better suited to understanding complex, contemporary risk problems because it acknowledges uncertainties rather than seeking to ignore them, and makes value judgments explicit.⁸¹ In their work Funtowicz and Ravetz sought to distinguish between three different types of scenarios concerning the use of science in policy-making, defined in terms of the interaction of two variables, 'systems uncertainties' and 'decision stakes'.⁸² 'Systems uncertainties' describes the intensity of uncertainty surrounding an issue where 'the problem is concerned not with the discovery of a particular fact, but with the comprehension or management of an inherently complex reality'.⁸³ 'Decision-stakes', on the other hand, connote 'all the various costs, benefits, and value commitments that are involved in the issue through the various stakeholders'.⁸⁴ Where both uncertainties and stakes are low (ie uncertainties can be managed using routine technical measures and processes, and stakes are simple and small) the resultant situation is characterized as one of 'applied science' where a scientific 'business as usual' approach works effectively.⁸⁵ A more complex situation arises in cases where one of either uncertainties or stakes is significant. Funtowicz and Ravetz characterize such decision-making as in the arena of 'professional consultancy', which involves the exercise of professional skill and judgment either to clarify the reliability of particular information or to resolve value questions.⁸⁶ In practice this can give rise to different opinions offered on the same hazard by different experts.

The final situation, in which both uncertainties and decision stakes are high, represents the sphere of post-normal science. When in the 'wild' area of post-normal science, Funtowicz and Ravetz contend that all (scientific and professional experts included) are 'amateurs' because the questions at issue are essentially 'trans-scientific' (that is, they can be asked of, but not answered by, science).⁸⁷ The authors' prescription for a 'quality assessment' of scientific materials in such circumstances is to make use of an 'extended peer community' that will use 'extended facts', including anecdotal and community

⁸⁰ Brian Wynne, 'Science and Social Responsibility' in Jake Ansell and Frank Wharton (eds), *Risk: Analysis, Assessment and Management* (John Wiley & Sons, Chichester, 1992) 142.

⁸¹ Silvio O Funtowicz and Jerome R Ravetz, 'Science for the Post-Normal Age' (1993) 25 *Futures* 740.

⁸² Silvio Funtowicz and Jerome Ravetz, 'Three Types of Risk Assessment and the Emergence of Post-Normal Science' in Sheldon Krimsky and Dominic Golding (eds), *Social Theories of Risk* (Praeger Publishers, Westport, CT, 1992) 253.

⁸³ Funtowicz and Ravetz (n 81) 744.

⁸⁵ *ibid* 745–7.

⁸⁷ Funtowicz and Ravetz (n 82) 253–4. The notion of trans-science draws on the work of Alvin Weinberg, 'Science and Trans-Science' (1972) 10 *Minerva* 209.

⁸⁴ *ibid*.

⁸⁶ *ibid* 749.

knowledge.⁸⁸ In effect, they see the problems of post-normal science as requiring a ‘democratization’ of science itself, not ‘out of some generalized wish for the greatest possible extension of democracy in society’ but rather because ‘an extension of peer communities, with the corresponding extension of facts, is necessary for the effectiveness of this new sort of science in meeting the great challenges of our age’.⁸⁹

A very useful distillation of these ideas, and their relevance for risk assessment, can be found in the work of Andy Stirling. Stirling’s research has focused particularly upon the precautionary principle and approaches for integrating this principle into risk assessment processes. In explaining the need for a precautionary risk-assessment approach to respond to certain kinds of uncertain, complex risk situations, Stirling explores different states of knowledge that may exist in any given situation about the two main parameters of conventional risk assessment, hazards/outcomes and likelihood/probabilities. Using these two parameters he differentiates between four logical permutations of possible states of incomplete knowledge,⁹⁰ which can be summarized as follows:

- *(true) Risk*: where there is high confidence about possible outcomes and their probabilities as assessors are dealing with familiar systems and controlled conditions (for example, the risk of meltdown of a reactor in a nuclear power plant).
- *Uncertainty*: where possible outcomes can be characterized but there is insufficient knowledge to assign probabilities because the systems under examination are complex, non-linear and open.
- *Ambiguity*: where it is the possible outcomes, rather than their probabilities, that are difficult to assess, whether because there is contestation over appropriate risk framings, assumptions and methods, experts disagree and/or issues of behaviour, trust or compliance are pertinent.
- *Ignorance*: where neither outcomes nor probabilities can be fully characterized such that effects may be unanticipated, there are gaps, surprises and unknowns, and/or the existence of novel causative mechanisms.⁹¹

Conventional processes of risk assessment, Stirling argues, offer ‘a powerful suite of methods under a strict condition of risk’ but are not ideal in risk situations characterized by uncertainty, ambiguity or ignorance.⁹²

⁸⁸ Funtowicz and Ravetz, *ibid* 254. See also the concept of ‘border work’ developed by Horlick-Jones and Sime (n 77) 445.

⁸⁹ Funtowicz and Ravetz, *ibid* 273.

⁹⁰ Andrew Stirling, ‘Risk at a Turning Point?’ (1999) 1 *J Environ Med* 119.

⁹¹ For a similar classification see Brian Wynne, ‘Uncertainty and Environmental Learning: Reconciling Science and Policy in the Preventative Paradigm’ (1992) 2 *Global Environ Change* 111.

⁹² Andrew Stirling, ‘Risk, Precaution and Science: Towards a More Constructive Policy Debate’ (2007) 8 *EMBO Reports* 311.

These insights from the social scientific literature on risk decision-making as to the circumstances in which standard scientific methods and conventional risk assessment yield inaccurate representations of risk, are also of relevance in considering the situations in which it is appropriate for a decision-maker reviewing a risk assessment or associated regulatory measures, to adopt a stringent, science-based standard of review.⁹³ If conventional notions of science and risk assessment are strictly applied in the review of measures adopted to deal with a situation of uncertainty, ambiguity or ignorance (or, in Funtowicz and Ravetz' framework, spheres of professional consultancy or post-normal science) this may privilege a narrow understanding of the nature of the risks at issue, which may not fully capture all the potential adverse outcomes of concern. Consequently, risk regulatory measures may be struck down as not being based on 'sufficient' scientific evidence or an 'adequate' risk assessment even though they are seeking to respond to a different vision of risk than that captured by conventional scientific methods.

A better calibration between the standard of review applied in SPS disputes and the circumstances in which scientific risk assessment (and hence science-based evaluation) holds its greatest validity could be achieved by allowing for greater flexibility in application of the standard of review, depending on the characteristics of the risk situation under consideration. The frameworks from the social scientific literature discussed above suggest two parameters of particular importance in differentiating between risk situations: the intensity of uncertainty (or incertitude)⁹⁴ surrounding a particular risk and the degree of socio-political contestation with respect to the issue that will give rise to different framings in risk assessment and different value judgments about the importance or otherwise of avoiding particular outcomes. In turn, differences over value orientations will tend to feed into disputes over the assumptions and methods to be applied in risk assessment (especially to deal with areas of incomplete knowledge). These two parameters, designated 'certainty' and 'consensus', were those put forward by a quintet of distinguished social science professors as a basis for distinguishing risk situations in an article published in the lead-up to the WTO panel's decision in the *Biotech* case.⁹⁵ The aim of David Winickoff and his co-authors in that article was to advocate for a procedural approach to SPS review, taking into account national values and policy judgments in scientific regulation.

⁹³ David Winickoff et al, 'Adjudicating the GM Food Wars: Science, Risk, and Democracy in World Trade Law' (2005) 30 *YaleJIntL* 81.

⁹⁴ Incertitude is often employed as a broader term to capture different states of incomplete knowledge including ambiguity, uncertainty and ignorance. However, in this article the term 'uncertainty', broadly construed, is preferred as it has been more commonly used in the legal literature and case law.

⁹⁵ Winickoff et al (n 93). The authors of the article were David Winickoff, Sheila Jasanoff, Brian Wynne, Lawrence Busch and Robin Grove-White.

<p>Uncertainty low</p> <p>Socio-political contestation high</p> <p><i>(Ambiguity, professional consultancy)</i></p>	<p>Uncertainty high</p> <p>Socio-political contestation high</p> <p><i>(Ignorance, post-normal science)</i></p>
<p>Uncertainty low</p> <p>Socio-political contestation low</p> <p><i>(Risk, scientific 'business as usual')</i></p>	<p>Uncertainty high</p> <p>Socio-political contestation low</p> <p><i>(Uncertainty, professional consultancy)</i></p>

Figure 1. Differentiating risk situations

Applying the approach of Winickoff et al more broadly to distinguish ‘apples from oranges’ so to speak, we can use combinations of the uncertainty and socio-political contestation parameters to identify four broad categories of risk situations that may confront decision-makers in an SPS (or other environmental) dispute (see [Figure 1](#)). It is acknowledged that, in the real world, the boundaries between each of these four risk situations will be permeable, which may require a decision-maker to judge which category best describes a particular risk situation (how panels and the Appellate Body might go about this task is discussed further in Part VI). It might also be argued that an approach characterizing risk situations according to two parameters only suffers from the same kind of reductionism that founds critiques of conventional risk assessment. While there is some truth to this criticism, the analysis of risk situations according to characteristics of uncertainty and socio-political contestation nonetheless offers a useful contribution in articulating categories of risk situations that fall outside the realm and expertise of conventional scientific risk assessment. Moreover, the parameters themselves are sufficiently broad to subsume a number of different elements commonly encountered in disputes over risk. For instance, the parameter of uncertainty captures problems of ignorance, differences between experts, inadequate methods and complex systems. On the other hand, the parameter of socio-political contestation brings into consideration different value assumptions, different risk framings and different assessments of the costs and benefits of a particular outcome.

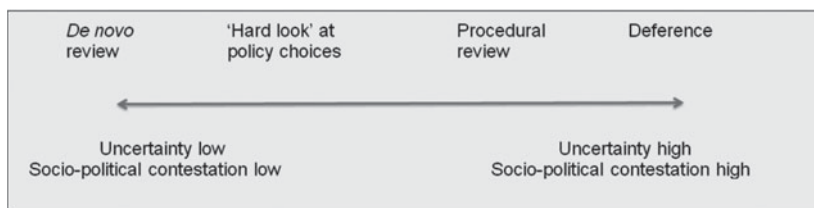


Figure 2. Standard of review continuum

The most difficult risk situations for assessment and decision-making purposes tend to be those where levels of uncertainty are high (perhaps because the risk is relatively novel (ie a case of ignorance) or research methodologies are not well-developed (leading to ambiguity)) and social consensus around risk framings is low (ie the extent and importance of the risk are perceived differently by different communities).⁹⁶ SPS disputes over risk situations falling into this category might justifiably attract a more deferential standard of review, both as there is more expert disagreement about what is the appropriate scientific basis for measures, and also because values that play a role in shaping the reasoning of risk assessors will be highly contested.

At the other end of the spectrum we would find risk situations where levels of uncertainty are low and there is a high degree of convergence in the social framing of the risks at issue (ie socio-political contestation over the risk is low). Here we would expect that standard scientific processes and conventional risk assessment would operate effectively and hence a fairly stringent standard of review would be justified to ensure that measures have a credible scientific underpinning and that risk assessment conclusions are adequately justified.

The two further categories of risk situations illustrated in Figure 1—low uncertainty/high socio-political contestation and high uncertainty/low socio-political contestation—will lie at intermediate points on the continuum of risk situations (see Figure 2). In situations of low uncertainty accompanied by high levels of socio-political contestation over the importance of the risk, a standard of review towards the stringent end of the continuum might be applied, similar to the ‘hard look’ approach taken by US federal courts. By contrast, risk situations involving SPS concerns characterized by high levels of uncertainty and also high levels of social consensus as to the importance of the risk (ie socio-political contestation is low) might warrant a more deferential standard of review, subject to procedural constraints such as those that have been employed by the European courts (eg independence of experts, quality of expertise, avenues for outside participation, transparency).

⁹⁶ Cass Sunstein describes these situations as ones involving ‘hot’ risks for which normative or cultural commitments play a central role in people’s assessments of risk: Cass R Sunstein, ‘Misfearing: A Reply’ (2006) 119 HarvLRev 1115.

Part VI takes up the question of how this approach to application of the standard of review might be operationalized in the practice of SPS dispute settlement. Before doing so, however, it is pertinent to consider the clarifications of the standard of review offered by the Appellate Body in its recent case law and whether they allow for the kind of differentiation between risk situations proposed.

IV. CONTINUED SUSPENSION (*HORMONES*)—A TURNING POINT?

On appeal of the *Continued Suspension* panel reports, one set of questions that faced the WTO Appellate Body concerned the appropriateness of the panel's approach in dealing with the scientific and expert evidence. Consistent with the de facto approach of *de novo* review that had been evolving in the SPS case law, the panel had essentially taken the majority view of its advising experts as a baseline for judging the adequacy of the EU's measures and their underlying scientific basis. The Appellate Body dealt with the issue of the appropriateness of the panel's approach in two ways: first, through an analysis of the requirements applicable to the selection and consultation of advising experts,⁹⁷ and secondly, and of most relevance for the present context, by focusing on the nature of the standard of review under the SPS Agreement. The Appellate Body's rulings on both questions demonstrated a strong appreciation of the important role that scientific evidence—and the experts who provide and evaluate that evidence—plays in SPS dispute settlement.

In respect of the standard of review applicable in SPS dispute settlement, the Appellate Body reiterated its previous findings in *Hormones* that the standard is one of 'objective assessment' derived from Article 11 of the DSU. However, it added several important clarifications with respect to what is required in order for a panel to make an objective assessment of a WTO Member's risk assessment. The Appellate Body ruled that it was not the task of the panel to determine the correctness of a Member's risk assessment but rather whether it 'is supported by coherent reasoning and respectable scientific evidence and is, in this sense, objectively justifiable'.⁹⁸ The Appellate Body then went on to detail the methodology panels should apply in reviewing a WTO Member's risk assessment, including the role of experts in this regard. The Appellate Body specified a four-step methodology consisting of the following stages:⁹⁹

1. *Identification of the scientific basis of the SPS measure.* The Appellate Body noted that this basis need not reflect a majority view within the scientific community but can embrace divergent or minority perspectives.

⁹⁷ For an analysis of the Appellate Body's rulings in this regard see Alessandra Arcuri, Lukasz Gruszczynski and Alexia Herwig, 'Independence of Experts and Standards for Evaluation of Scientific Evidence under the SPS Agreement – New Directions in the SPS Case Law' (2010) 2 European Journal of Risk Regulation 183 and Cho (n 6) 302.

⁹⁸ *Continued Suspension*, para 590.

⁹⁹ *ibid* para 591.

2. *Verification that the scientific basis comes from a respected and qualified source.* Accordingly, the Appellate Body ruled that the scientific material relied upon in risk assessment must have the ‘necessary scientific and methodological rigour to be considered reputable science’. This does not imply general acceptance of a particular view within the scientific community but rather evidence ‘considered to be legitimate science according to the standards of the relevant scientific community’.
3. *Assessing whether the reasoning articulated on the basis of the scientific evidence is objective and coherent.* To meet this standard the Appellate Body held that the conclusions drawn by a Member need to find sufficient support in the scientific evidence relied upon.
4. *Determining whether the results of the risk assessment sufficiently warrant the SPS measure.* According to the Appellate Body, the task at this stage was to discern whether there is a rational or objective relationship between the scientific basis identified and the SPS measure adopted, although the scientific basis can be a minority view if coming from a qualified and respected source.

The role of the experts advising the panel is to assist the panel with the above tasks while avoiding the situation of *de novo* review where the panel asks whether ‘the experts would have done a risk assessment in the same way and would have reached the same conclusions as the risk assessors’.¹⁰⁰

On its face, the methodology articulated by the WTO Appellate Body in *Continued Suspension* would seem to allow greater scope for deferential review of a Member’s challenged risk assessment in SPS dispute settlement. This is the view of commentators such as Gruszczynski, who considers that the ‘standard is much closer to the deferential end of the spectrum’ although ‘it would be too far to conclude that it is a fully procedural one’.¹⁰¹ He concludes that the new standard

leaves WTO Members with a great degree of discretion as to how to assess scientific data and what kind of inferences to make on this basis. It also greatly reduces the need of the panel to engage in a detailed examination of scientific evidence and deciding which scientific view is better.¹⁰²

Accepting Gruszczynski’s analysis, we might expect that the standard of review articulated in *Continued Suspension* would allow greater scope for the adoption of legitimate SPS measures in categories of risk situations exhibiting characteristics of scientific uncertainty and some degree of socio-political contestation.

There are certainly a number of indicators in the Appellate Body’s judgment in *Continued Suspension* that suggest it favoured a more deferential approach to evaluating risk assessment and the supporting scientific evidence than had

¹⁰⁰ *ibid* para 592.

¹⁰¹ Gruszczynski (n 6) 146.

¹⁰² *ibid*.

become the practice in SPS jurisprudence. In particular, in respect of assessing the ‘scientific basis’ of an SPS measure, the Appellate Body directed panels to focus on the respectability and reputability of the *source* of the scientific evidence rather than evaluating the cogency of the scientific evidence itself. This approach has some similarities with that taken by the US Supreme Court in the well-known *Daubert* case dealing with the question of which evidence qualifies as expert testimony under federal evidence rules.¹⁰³ In its decision, the Supreme Court relied on a flexible set of factors designed to ensure that admitted scientific evidence is both relevant and reliable. Since the Supreme Court’s decision, however, the *Daubert* test has proven no easier for courts to apply in practice than the previous standard,¹⁰⁴ which relied on assessing whether a particular view or technique had gained ‘general acceptance’ in the particular scientific field to which it belonged.¹⁰⁵ Moreover, the Supreme Court’s emphasis on the ‘scientific validity’ of evidence has tended to push reviewing courts to impose stricter requirements on the admission of expert opinion.

The US experience under *Daubert* suggests the need for caution before hailing the *Continued Suspension* decision as ‘a turning point’ with respect to the standard of review.¹⁰⁶ In particular, the Appellate Body’s review methodology outlined in *Continued Suspension* still contains a substantial emphasis on scientific factors (eg the ‘scientific basis’ for measures, which must display ‘scientific and methodological rigour’ sufficient to be considered ‘legitimate science’ by reference to ‘the standards of the relevant scientific community’).¹⁰⁷ Although panels are instructed not to disregard scientific evidence that comes from a minority of scientists, they will still be relying heavily on their advising experts to help them identify what evidence should or should not be considered ‘legitimate science’, an evaluation which may well gravitate towards peer reviewed studies or data that has widespread acceptance in the relevant scientific community. It is also worth noting that the Appellate Body cast its review methodology in the context of evaluation of risk assessment and so it may not be applicable to assessments of the ‘sufficiency’ or ‘insufficiency’ of scientific evidence under other provisions of the SPS Agreement. It has frequently been in assessing this element of SPS measures that the heaviest scrutiny of the disputed science has been evident.

¹⁰³ *Daubert v Merrell Dow Pharmaceuticals*, 113 S Ct 2786 (1993).

¹⁰⁴ Michael Saks and David Faigman, ‘Expert Evidence After Daubert’ (2005) 1 Annual Review of Law and Social Science 105.

¹⁰⁵ *Frye v. United States*, 54 App. D.C. 46, 47, 293 F. 1013, 1014 (1923).

¹⁰⁶ Gruszczynski (n 6) 52.

¹⁰⁷ WorldTradeLaw.net LLC, *Appellate Body Reports: Canada/United States – Continued Suspension of Obligations in the EC – Hormones Dispute* (2008) 30.

V. AUSTRALIA—APPLES—A DOUBLE STANDARD?

The *Australia–Apples* case offered the first opportunity for application of the new standard of review requirements in WTO SPS dispute settlement. Like the previous *Japan–Apples* case, the *Australian–Apples* dispute involved quarantine measures designed to prevent the introduction of various pests and diseases, including fire blight, into the territory of the Member concerned (Australia) from other countries where these diseases and pests were established. New Zealand challenged the Australian measures arguing that there was ‘no scientific support’ for Australia’s contention that imported apples are a pathway for transmitting various plant diseases and pests, and hence that the Australian measures were inconsistent with Article 2.2 of the SPS Agreement. New Zealand also argued that the Import Risk Analysis (IRA) produced by Australian authorities to justify and elaborate quarantine restrictions on New Zealand apples was inconsistent with the risk assessment requirements under Article 5.1 of the SPS Agreement. These arguments were upheld by at first instance by the panel and on appeal to the Appellate Body.

This result might have seemed inevitable given the *Japan–Apples* precedent, however, the risk assessment put forward by the Australian quarantine authorities differed significantly from that struck down as inadequate in the Japanese case. The Australian IRA adopted what was termed a ‘semi-quantitative’ approach that combined a quantitative estimation of the probability of entry, establishment and spread of particular diseases or pests with a qualitative evaluation of the economic and environmental consequences.¹⁰⁸ This approach was adopted because of the concerns expressed by stakeholders during the domestic decision-making process over the transparency and objectivity of using a purely qualitative risk assessment.

The quantitative estimates of probabilities made by the IRA were expressed as probability intervals, rather than as single numbers. For instance, a probability interval of zero to 10^{-6} (1 in a million) was used for a negligible likelihood of occurrence of a particular event. The IRA also employed probability distributions to pinpoint a value within each interval on a per apple basis for each importation step of an importation scenario and the different pathways of distribution, utilization, waste generation and disposal of apples in Australia. Where insufficient information was available to determine a most likely value within a given probability interval, the IRA applied a uniform distribution approach on the basis of ‘expert judgment’. A uniform distribution model assumes that each value in the probability interval range occurs with equal probability. In the case of a negligible likelihood with a probability interval of $0-10^{-6}$, application of a uniform distribution rule yields a midpoint of 0.5 in a million.

¹⁰⁸ Biosecurity Australia (2006) *Final import risk analysis report for apples from New Zealand*, Part A, Biosecurity Australia, Canberra.

This IRA methodology was heavily criticized by the panel in *Australia–Apples*. The panel focused particularly on the choice of a probability interval of 0 to 10^{-6} and use of uniform distribution to model events with a ‘negligible’ likelihood of occurring, ruling that these choices were not properly justified in the IRA and led to an overestimation of the probability of entry, establishment and spread of the diseases and pests at issue. The panel identified these ‘flaws’ as ‘magnify[ing] the assessment of risk, turning what are often the remotest of possibilities into events that are assessed as occurring with some frequency.’¹⁰⁹ This finding formed a major plank of the panel’s overall ruling that the IRA was not a proper SPS risk assessment. Arguably the panel’s finding came very close to saying that it disagreed with the risk assessment methodology chosen by the WTO Member concerned and that it (or its advising experts) would have done the risk assessment differently.

Consequently, on appeal, a central part of Australia’s arguments was that the panel had misinterpreted and misapplied the standard of review articulated in *Continued Suspension* in its assessment of the IRA under Article 5.1.¹¹⁰ However, the Appellate Body upheld the panel’s application of the standard of review and its finding that the IRA’s conclusions were not objective or coherent because they exaggerated or overestimated certain risks and consequences and did not find sufficient support in the scientific evidence relied upon. In this regard, the Appellate Body drew a distinction between the scrutiny to be applied to the underlying *scientific basis* of a measure, and scrutiny of the *reasoning* of a risk assessor on the basis of the scientific evidence. In the case of the former, the standard to be applied was more deferential reflecting the fact that panels are not well-equipped to undertake their own scientific assessment.¹¹¹ Accordingly, the applicable standard was whether the scientific material constitutes ‘legitimate science according to the standards of the relevant scientific community’.¹¹²

By contrast, when reviewing the reasoning of risk assessors, the apparently more stringent standard of objectivity, coherence and assessment of the sufficiency of scientific evidentiary support applied.¹¹³ The Appellate Body found the IRA’s reasoning and conclusions drawn in the risk assessment fell into this second category, even where expert judgment was exercised by risk assessors to address alleged scientific uncertainty. Indeed, the Appellate Body was of the view that ‘when the exercise of expert judgment forms an integral part of the risk assessor’s analysis, then it should be subject to the same type of [more stringent] scrutiny by the panel as all other reasoning and conclusions in the risk analysis.’¹¹⁴ The Appellate Body did not view the phrase ‘as appropriate to the circumstances’ in Article 5.1 of the SPS Agreement as

¹⁰⁹ *Australia – Measures Affecting the Importation of Apples from New Zealand*, Report of the Panel, WT/DS367/R, 9 August 2010, para 7.508, quoting the New Zealand submission.

¹¹⁰ *Australia–Apples*, para 216.

¹¹¹ *ibid* para 225.

¹¹³ *Australia–Apples*, para 224.

¹¹² *Continued Suspension*, para 591.

¹¹⁴ *ibid* paras 222–4.

affording such flexibility as to excuse a risk assessor from properly performing the risk assessment.¹¹⁵ Hence although recourse to expert judgement in the IRA was not in itself objectionable, ‘it must be reasoned and explained consistently with Articles 5.1 and 5.2 of the SPS Agreement so that the risk assessment can still be considered a scientific process that is based on the “available scientific evidence”.’¹¹⁶

The Appellate Body’s findings in *Australia–Apples* raise an important question about the current interpretation of the SPS standard of review, which has been succinctly summarized by Gruszczynski:

One may consider *US/Canada–Continued Suspension* as an anomaly in an otherwise rather consistent line of cases that subscribed to *de novo* review. However, it is also possible that the Appellate Body wants to apply a more deferential standard of review in human health related trade disputes (such as *Continued Suspension*), while in traditional phytosanitary cases the applicable standard will remain intrusive.¹¹⁷

If the Appellate Body were indeed seeking to distinguish between different types of SPS disputes in terms of the stringency of the review applicable, what would be the basis for that distinction? As discussed in Part II, the Appellate Body has relied in the past on the text of the SPS Agreement to distinguish between two categories of risk situation: 1) food or feed safety risks where it is only necessary in risk assessment to evaluate *possible* adverse effects on human or animal health; and 2) quarantine risks associated with the introduction of pests or diseases via trade in a particular commodity in which case an evaluation of the *probability* of harm is required.

At a normative level, however, there is no compelling rationale for distinguishing, on this basis, the applicable risk assessment standard and associated stringency of judicial review. For one, both categories of risk (food-borne and quarantine) may relate to human health,¹¹⁸ so that if there was some desire to elevate the value of human health protection above other regulatory purposes it is not served by the Appellate Body’s textual interpretation. Even if the intention was to require a higher standard of justification for measures addressing quarantine risks, the Appellate Body’s tolerance of both qualitative and quantitative forms of risk assessment blurs the distinction where low likelihood risks are at issue. As Alessandra Arcuri has astutely noted, there may be little difference in practice between assessing whether an event is possible and assessing whether it is of low or negligible probability.¹¹⁹ The difficulties

¹¹⁵ *ibid* para 237.

¹¹⁶ *ibid* para 241.

¹¹⁷ Lukasz Gruszczynski, ‘How Deep Should We Go? – Searching for an Appropriate Standard of Review in the SPS Cases’ (2011) 1 *European Journal of Risk Regulation* 58.

¹¹⁸ Lukasz Gruszczynski, ‘Science in the Process of Risk Regulation under the WTO Agreement on Sanitary and Phytosanitary Measures’ (2006) 7 *German Law Journal* 382.

¹¹⁹ Alessandra Arcuri, ‘Food Safety at the WTO after Continued Suspension: a Paradigm Shift?’ in Antonis Antoniadis, Robert Schütze and Eleanor Spaventa (eds), *The European Union and Global Emergencies: A Law and Policy Analysis* (Hart Publishing, London, 2011) 205.

experienced with the IRA in *Australia–Apples* bear this out. One can speculate that the IRA may have fared better in WTO dispute settlement had the Australian quarantine authorities opted for a purely qualitative (but less transparent) assessment of the probabilities of introduction, establishment and spread of plant pests and diseases via apple imports.

Arcuri seeks to explain the different approaches evident in the SPS case law as the result of the Appellate Body ‘juggling’ between two different visions of risk put forward by competing clusters of experts and scholars in the field of risk analysis. The first cluster, which she says subscribes to a ‘quantitative-risk logic’, adopts the conventional technical risk perspective in which risk is a calculable entity assessed via scientific means. The second group, falling into the constructivist school, employs a ‘holistic-risk logic’ that recognizes the role of values and socio-cultural context in the process of risk assessment. Arcuri hypothesizes that as WTO panels and the Appellate Body ‘juggle’ between these two visions of risk in SPS disputes they end up favouring different standards of risk assessment and review.¹²⁰

VI. OPERATIONALIZING A DIFFERENTIATED STANDARD OF REVIEW

While Arcuri’s juggling hypothesis provides a useful analytical lens for explaining the changing stringency of review applied in different SPS disputes, continued ‘juggling’ of this kind by the Appellate Body does not provide a sustainable, principled solution for application of the standard of review in SPS dispute settlement. Moreover, implicit in the Appellate Body’s ‘juggling’ between different stringencies of review is the suggestion that there exists a hierarchy of SPS risks. It is striking that the cases where the Appellate Body has adopted a more deferential stance have been those involving human health risks (cancer potentially caused by residues of growth hormones in beef), whereas the standard of review has uniformly been more stringently applied in phytosanitary quarantine cases.

Some might argue that the solution to the standard of review question therefore lies in the Appellate Body and panels making explicit the value judgments they are exercising regarding the importance of different SPS risks.¹²¹ There is some precedent for this approach in a number of cases decided by the Appellate Body under the General Agreement on Tariffs and Trade (GATT) involving health and environmental risks. In its jurisprudence on the ‘necessity’ test applicable under the GATT health exception, Article XX (b), the Appellate Body has sought to ‘weigh and balance’ a series of factors

¹²⁰ Alessandra Arcuri, ‘Interpreting the Concepts of “Risk Management” and “Insufficiency of Scientific Evidence”’: Juggling between the Logics of Different Knowledge-based Groups?’ in Maria-Chiara Malaguti et al (eds), *Science and Law – Scientific Evidence in International and European Law* (Argo Editore, Rome, 2009) 66.

¹²¹ M Gregg Bloche, ‘WTO Deference to National Health Policy: Toward an Interpretive Principle’ (2002) 5 JIEL 825.

which include ‘the importance of the common interests or values protected’ by a particular measure.¹²² In the case of *Asbestos*, concerning the GATT legality of a health-related ban on asbestos products, the Appellate Body gave great weight to the health objective pursued by the measure in evaluating the necessity for its adoption. It described the ‘value’ of ‘the preservation of human life and health through the elimination, or reduction, of the well-known, and life-threatening health risks posed by asbestos fibres’ as ‘both vital and important in the highest degree’.¹²³ Similarly, in the later case of *Retreaded Tyres*, which also involved arguments under Article XX(b), both the panel and the Appellate Body characterized the objective of environmental protection pursued by the disputed measure as important.¹²⁴ Given the importance of the interests protected by the import ban (which also included health objectives), the ban was judged necessary despite its trade-restrictiveness.¹²⁵

There are a number of reasons, however, to be cautious about transferring this approach to the SPS context as a way of supporting the application of different standards of review in different cases. One immediate concern that arises is how international decision-makers involved in SPS dispute settlement might legitimately go about evaluating the importance of different SPS risks and the values attached to them. Even in the national sphere there are few established guidelines for determining the relative substantive importance of interests or values in the health or environmental sphere.¹²⁶ In addition, SPS measures have a particular focus on risks arising *within* a Member’s territory making questions of local environmental conditions and the risk preferences of local populations more pertinent.¹²⁷ Universal value commitments are thus difficult to discern in these circumstances.

It also remains unclear from the GATT health and environmental jurisprudence to date whether an approach that looks to the importance of the values protected by a measure would lead to a more deferential review

¹²² *Korea – Measures Affecting Imports of Fresh, Chilled and Frozen Beef*, Report of the WTO Appellate Body, WT/DS169/AB/R, 11 December 2000, para 164. In the context of article XIV of the General Agreement on Trade in Services, which also includes a necessity test, the Appellate Body indicated in *United States – Measures Affecting the Cross-Border Supply of Gambling and Betting Services*, Report of the WTO Appellate Body, WT/DS285/AB/R, 7 April 2005 that the weighing and balancing process ‘begins with an assessment of the “relative importance” of the interests or values furthered by the challenged measure’: para 306.

¹²³ *European Communities – Measures Affecting Asbestos and Asbestos-Containing Products*, Report of the WTO Appellate Body, WT/DS135/AB/R, 12 March 2001 (*Asbestos*), para 172.

¹²⁴ *Brazil – Measures Affecting Imports of Retreaded Tyres*, WT/DS332/R, 12 June 2007 (Report of the Panel), para 7.112; *Brazil – Measures Affecting Imports of Retreaded Tyres*, Report of the Appellate Body, WT/DS332/AB/R, 3 December 2007 (*Retreaded Tyres*), para 179.

¹²⁵ *Retreaded Tyres*, para 179.

¹²⁶ Timothy Reif and Julie Eckert, ‘Courage You Can’t Understand: How to Achieve the Right Balance Between Shaping and Policing Commerce in Disputes Before the World Trade Organization’ (2004) 42 *ColumbJTransnatlLaw* 689.

¹²⁷ Consequently Article 5.2 of the SPS Agreement provides that in risk assessment Members shall take into account, *inter alia*, ‘prevalence of specific diseases and pests; existence of pest- or disease-free areas; relevant ecological and environmental conditions’.

approach where the risks at issue extend beyond situations of well-founded harms. In *Asbestos*, the willingness of the Appellate Body to find the values at stake were vital seemed to be predicated on a finding of scientifically well-supported health risk.¹²⁸ In other words, it appeared to be the strong scientific consensus surrounding the existence of asbestos-related health risks that founded the Appellate Body's rulings that there was also substantial social consensus on the importance of the interests at stake. Faced with a situation where disputes remain over the scientific basis of risk concerns (the issue of genetically modified organisms is a good example here), WTO decision-makers might be more hesitant to apply a deferential standard of review on the basis of value concerns alone.¹²⁹

This article has argued that a better approach to recognizing the diversity of risk situations presented by SPS cases, and potentially also other international environmental disputes, would be to differentiate between risk situations according to their characteristics of uncertainty and degree of socio-political contestation, and to tailor the stringency of the standard of review applied accordingly. The parameters of uncertainty and socio-political contestation offer the potential for a more transparent, reasoned assessment process than in the case of different values. As regards the criterion of uncertainty there are now numerous analyses that describe the different types of uncertainty encountered in scientific research and put forward indicia for recognizing whether uncertainty arises in a particular situation.¹³⁰ Discerning differing degrees of socio-political contestation regarding risks may be more problematic. However, Winickoff et al suggest one approach would be to look to evidence sourced in the risk literature, regulatory experience or public dialogues which suggests a lack of consensus as to the nature, sources and extent of the risks involved.¹³¹ Using such evidence, the task of a decision-maker would be to assign a ranking to each parameter of 'high' or 'low'. The combination of the parameters assigned (eg high/high, low/high etc) could then guide the applicable stringency of review, as illustrated in [Figure 1](#).

While in theory this approach seems straightforward more difficult issues of interpretation are likely to be encountered in SPS dispute settlement practice. Concerns already exist that legally-trained decision-makers such as WTO panelists are ill-equipped to grapple with scientific and technical evidence.¹³² While the Appellate Body in a number of its decisions has demonstrated considerable sophistication in its understanding of different perspectives

¹²⁸ Button (n 12) 4.

¹²⁹ Claims under the GATT were raised by Canada and Argentina but not addressed by the panel in the *Biotech* case.

¹³⁰ For a legal audience a useful typology is found in Vern Walker, 'The Siren Songs of Science: Toward a Taxonomy of Scientific Uncertainty for Decisionmakers' (1991) 23 *Connecticut Law Review* 567.

¹³¹ Winickoff et al (n 93) 116.

¹³² Pauwelyn (n 25).

on science and risk assessment,¹³³ it is panels who continue to bear the brunt of the task of evaluating the factual information including the ever-growing volume of scientific and technical evidence in SPS disputes.¹³⁴ Would they fare any better if asked to discern levels of uncertainty pertaining to a risk and the degree to which there are different risk framings, value judgments and assumptions at stake? Perhaps not. However, recognizing these matters as relevant factors for assessment by a panel might at least permit the involvement of a broader range of perspectives and evidence about the risks under consideration, potentially also extending beyond the views of experts to include anecdotal and regulatory experience, and public opinion.

Another way of approaching the problem, familiar to lawyers, is to reason by analogy from existing examples. If we go back over the past SPS case law (and this could also be done for disputed issues that have come before the WTO SPS committee) we might categorize the types of risk situations that have arisen as shown in Figure 3 below. In evaluating the characteristics of a risk situation in a later SPS dispute, decision-makers might consider whether that situation is more or less like those that have been encountered in the past.¹³⁵

Finally, decision-makers might make use of other tools in combination with differentiations in the stringency of review to cater for the particular characteristics of different risk situations. This may be especially useful in cases that occur at the boundary between different categories of risk situations. For instance, in situations of low to moderate uncertainty but high socio-political contestation, a key issue will most likely be that parties dispute the range of possible adverse outcomes that may occur. In such cases, a relatively stringent, ‘hard look’ style of review might be coupled with a proportionality-style analysis that looks to the coherence of the relationship between the assessed level of risk and the SPS measures adopted in response.¹³⁶ Cost–benefit analysis could potentially also be a useful tool in this instance in clarifying the financial costs and social benefits associated with

¹³³ See particularly the Appellate Bodies’ reports in *Hormones* and *Continued Suspension*. On the question of the level of scientific uncertainty (or insufficiency of the scientific evidence in the terminology of the SPS Agreement), the Appellate Body in *Continued Suspension* demonstrated how this concept can be understood as a spectrum running from well-settled scientific theories to situations where new theories generate a ‘paradigm shift’ in prevalent scientific thinking: para 703.

¹³⁴ This evidence amounted to hundreds of pages in the *Biotech* case: see para 739 and Annexes H–J.

¹³⁵ This is a somewhat familiar analysis given the requirements of Article 5.5 of the SPS Agreement. For a note of caution with respect to the comparability of risk situations see Jeffery Atik, ‘The Weakest Link: Demonstrating the Inconsistency of “Appropriate Levels of Protection” in *Australia–Salmon*’ (2004) 24 Risk Analysis 483.

¹³⁶ Articles 5.5 (requiring consistency in the treatment of similar risk situations) and Article 5.6 (requiring that SPS measures adopted to achieve a Member’s ALOP are not more trade-restrictive than required) could provide the legal underpinning for such an approach.

<p>Uncertainty low</p> <p>Socio-political contestation high</p> <p><i>(Ambiguity, professional consultancy)</i></p> <ul style="list-style-type: none"> • Hormone residues in beef <p><i>(Hormones, Continued Suspension).</i></p>	<p>Uncertainty high</p> <p>Socio-political contestation high</p> <p><i>(Ignorance, post-normal science)</i></p> <ul style="list-style-type: none"> • Genetically modified organisms (<i>Biotech</i>). • Mad cow disease transmission to humans in the early days of appearance of this risk.
<p>Uncertainty low</p> <p>Socio-political contestation low</p> <p><i>(Risk, scientific ‘business as usual’)</i></p> <ul style="list-style-type: none"> • Necessity for testing the efficacy of chemical quarantine treatment for different fruit varieties <p><i>(Varietals).</i></p> <ul style="list-style-type: none"> • Phytosanitary risks associated with importing diseases and pests on fruit <p><i>(Apples).</i></p>	<p>Uncertainty high</p> <p>Socio-political contestation low</p> <p><i>(Uncertainty, professional consultancy)</i></p> <ul style="list-style-type: none"> • Importation of fish diseases in imported salmon (<i>Salmon</i>).

Figure 3. Differentiating risk situations in the SPS case law

different regulatory options that respond to different understandings of the potential hazards.¹³⁷

Conversely, in risk situations involving SPS concerns characterized by high levels of scientific uncertainty but low to moderate levels of socio-political contestation (ie the risk is generally regarded as being important and framings converge) it may be that uncertainties will be able to be reduced over time. In such cases, a more deferential standard of review that focuses on the adequacy of procedures adopted by the Member in assessing risks (ie their openness, transparency, use of qualified expertise, etc) might be supplemented by the acceptance of measures based on that assessment as a precautionary response applying on a provisional basis only.¹³⁸

¹³⁷ See further, Jeremy D Fraiberg and Michael J Trebilcock, ‘Risk Regulation: Technocratic and Democratic Tools for Regulatory Reform’ (1998) 43 McGillLJ 835.

¹³⁸ Scope for such measures is offered by Article 5.7 of the SPS Agreement.

The approach described above, based upon differentiating different SPS risk situations and the stringency of the standard of review applied, is unlikely to satisfy all stakeholders: some will feel that it still allows too much scope for protectionism; others that it allows insufficient scope for including public risk perceptions. It may also be seen as introducing unacceptable levels of discretion into the task of international adjudication.¹³⁹ However, judicial review of science-based risk regulation will always remain a normative exercise at some level.¹⁴⁰ Arguably, a methodology for adjusting the stringency of the standard of review according to qualities of the risk situation at hand provides a more principled and transparent basis for the science-law interaction in SPS disputes than the current ‘juggling’ evident in Appellate Body decision-making. It also has the advantage that it can be relatively easily accommodated within the existing provisions and jurisprudence under the SPS Agreement, for example, by relying on the criterion of sufficiency/insufficiency of scientific evidence in Articles 2.2 and 5.7, as well as the Appellate Body’s recognition in *Hormones* of the necessity for the inclusion of ‘real world’ considerations in SPS risk assessment.

VII. CONCLUSION

This article has examined a seemingly technical question—the standard of review applicable in WTO SPS disputes—but in so doing has sought to demonstrate the broader implications of this issue for domestic risk regulation, and questions surrounding the use of science and expertise in WTO dispute settlement. The Appellate Body case law on the SPS standard of review has flip-flopped between deferential and stringent applications, but without articulating a clear normative rationale for this approach. In this article an alternative approach has been put forward, based upon differentiating SPS risk situations according to key parameters of associated levels of scientific uncertainty and socio-political contestation around risks, and adjusting the stringency of judicial review applied to reflect different combinations of these two parameters. It is argued that this manner of approaching the standard of review in SPS cases would allow for a more coherent and principled application in practice that recognizes the diversity of risk situations falling within the scope of the SPS Agreement.

In the first WTO dispute to come before the Appellate Body, it declared that WTO law should not be interpreted ‘in clinical isolation’ from the rest of public international law.¹⁴¹ Equally it should not be assumed that developments occurring in WTO SPS law are of no relevance to the broader sphere of international dispute settlement. Highly technical disputes, raising difficult

¹³⁹ Foster (n 26) 16–17.

¹⁴⁰ Peel (n 58).

¹⁴¹ *United States – Standards for Reformulated and Conventional Gasoline*, Report of the WTO Appellate Body, WT/DS2/AB/R, 29 April 1996.

questions over the methodology to be applied by international courts and tribunals in evaluating scientific evidence, are increasingly a feature of international environmental law.¹⁴² As the *Pulp Mill* case indicates, other international judicial tribunals may look to WTO SPS dispute settlement as a site of international ‘best practice’ on questions relating to the appropriate treatment and use of science and expertise in decision-making. To the extent that international legal tribunals dealing with complex, fact-intensive environmental disputes seek to model their own practices on that of WTO SPS dispute settlement through the resort to independent expert advice and science-based standards of evaluation, they are likely to face similar challenges and questions about how experts are selected, how their independence is determined, how much weight is placed on the experts’ views, and ultimately, the appropriate standard of review. There is thus considerable potential for innovations in the application of the standard of review in SPS dispute settlement to provide a useful model for environmental dispute resolution in international adjudicatory fora beyond the realm of the WTO.

¹⁴² Pertinent examples include *Methanex Corporation v United States of America*, NAFTA Chapter 11 Arbitral Tribunal (2005) 44 ILM 1345, Part IIIA, paras 37–40 and *Responsibilities and Obligations of States Sponsoring Persons and Entities with Respect to Activities in the Area* (Advisory Opinion) (Seabed Disputes Chamber of the International Tribunal for the Law of the Sea, ITLOS Case No 17, 1 February 2011), para 131.