

# Risk and Regulatory Calibration: WTO Compliance Review of the US Dolphin–Safe Tuna Labeling Regime

CARY COGLIANESE \*

*University of Pennsylvania Law School*

ANDRÉ SAPIR \*\*

*Université Libre de Bruxelles (ECARES), Bruegel and CEPR*

**Abstract:** In a series of recent disputes arising under the TBT Agreement, the Appellate Body has interpreted Article 2.1 to provide that discriminatory and trade-distortive regulation could be permissible if based upon a ‘legitimate regulatory distinction’. In its recent compliance decision in the *US–Tuna II* dispute, the AB reaffirmed its view that regulatory distinctions embedded in the US dolphin-safe tuna-labeling regime were not legitimate because they were not sufficiently calibrated to the risks to dolphins associated with different tuna fishing conditions. This paper analyzes the AB’s application of the notion of risk-based regulation in the *US–Tuna II* dispute and finds the AB’s reasoning lacking in coherence. Although risk analysis and calibration can in principle play useful roles in TBT cases, the AB needs to provide more explicit and careful guidance to WTO members and to panels to avoid the kind of ad hoc decision-making exhibited throughout the *US–Tuna II* dispute.

We live in a world filled with risk, and risk has come to dominate discourse over a wide range of domestic and international policy concerns (Beck, 1992; Jasanoff, 1999; Boyd, 2012). Risk obviously underlies multilateral environmental treaties related to trade,<sup>1</sup> and it finds expression in specific trade agreements on the use of risk assessment to guide domestic policymaking, such as the World Trade Organization (WTO) Agreement on Sanitary and Phytosanitary Measures (Howse, 2000). Risk has even found its way into general trade law principles. Article 2.2 of the WTO Agreement on Technical Barriers to Trade (TBT), for

\* Email: [cary\\_coglianese@law.upenn.edu](mailto:cary_coglianese@law.upenn.edu)

\*\* Email: [andre.sapir@ulb.ac.be](mailto:andre.sapir@ulb.ac.be)

The first author would like to thank Paul Stephan for his research assistance. The second author would like to thank the first author for their collaboration and emphasize that the first author is listed first not for alphabetical reasons but because he is actually the main author. Both authors convey their gratitude for helpful comments from Adam Finkel, Petros Mavroidis, Boris Rigod, David Zaring, and participants in the conference on The Legal–Economic Analysis of WTO Case Law of 2015.

<sup>1</sup> An example would be the Basel Convention on transboundary shipments of hazardous wastes.

example, calls for trade-restrictive domestic regulations to be necessary to meet one or more legitimate objectives, such as the protection of human health and safety – objectives which Article 2.2 explicitly refers to as ‘risks’ that demand ‘assessing’.

In this article, we consider a recent trade dispute resolved principally under Article 2.1 of the TBT to draw out lessons about the treatment of risk in international trade law. Although this anti-discrimination provision makes no explicit mention of risk, we show how risk concepts figured centrally in its application, as exemplified in one of the most recent WTO Appellate Body (AB) decisions to examine this provision: its compliance decision in the *US–Tuna II* dispute.<sup>2</sup> This decision avoided raising many major new questions about issues previously dissected by trade scholars, such as what counts as a ‘technical regulation’ or a ‘like product’ (e.g., Mavroidis, 2013). Instead, it centered on whether a US regulatory regime that impacted Mexico in adverse and differential ways was nevertheless justified on the grounds of a ‘legitimate regulatory distinction’. Central to that determination was the AB’s explicit risk-based approach, focusing on how well ‘calibrated’ the US regime was to the risks of harm to dolphins from tuna fishing. The AB found that the US measure was not properly calibrated.

The risk-based orientation adopted by the Appellate Body in *US–Tuna II* was hardly novel, nor is it necessarily unreasonable in principle. But a closer analysis of the treatment of risk and regulatory calibration throughout this proceeding suggests a need for more careful and coherent guidance from the Appellate Body in the future as to the principles that states should follow in calibrating risks and regulatory distinctions – and principles that panels should use in adjudicating future TBT disputes. Although reliance on the construct of risk may make legal and policy analysis appear more objective and rigorous, we caution that, without careful articulation of underlying principles, the invocation of risk in the resolution of disputes about trade discrimination may ironically contribute to ad hoc or incoherent decision-making.

## 1. The long legal dispute over dolphin-safe tuna

Trade disputes involving US–dolphin protection laws date back at least to the 1990s.<sup>3</sup> Although the earliest disputes centered on US embargoes of tuna and tuna products caught in ways that could not be shown to be ‘dolphin-safe’, the most recent series of WTO decisions stem from a US regulatory regime governing the labeling of tuna and tuna products as dolphin-safe. This US labeling regime is built upon the 1990 Dolphin Protection Consumer Information Act (DPCIA)

<sup>2</sup> Article 21.5 Appellate Body Report, *United States – Measures Concerning the Importation, Marketing and Sale of Tuna and Tuna Products*, WT/DS381/AB/RW, 20 November 2015 (Compliance (Art. 21.5)).

<sup>3</sup> For a historical account beginning in the 1970s, see Parker (1999).

(amended in 1997), and includes regulations<sup>4</sup> implementing the statute and a decision of the US Court of Appeals for the Ninth Circuit arising under the DPCIA.<sup>5</sup> These laws, among other things, preclude labeling as dolphin-safe any tuna caught in the Eastern Tropical Pacific Ocean (ETP) through nets that encircle tuna swimming with dolphins – that is, by ‘setting on’ dolphins.

Mexico initiated a trade action against these regulations in 2008 because most of its tuna fishing fleet operates within the ETP using methods that encircled tuna swimming with dolphins. A WTO Panel found in 2011 that the US regulations did not offend Article 2.1 of the Technical Barriers to Trade (TBT) Agreement but they did violate Article 2.2 by being ‘more trade-restrictive than necessary to fulfill their legitimate objectives’.<sup>6</sup>

In 2012, the Appellate Body issued a decision reviewing the original Panel Report. Although the AB still held for Mexico, it effectively flipped and found that the US had violated Article 2.1 but not Article 2.2.<sup>7</sup> That original AB decision found that the US labeling requirements were less favorable to Mexico in their differential treatment of tuna fished by Mexican boats in the ETP compared with tuna from elsewhere, and it held that the difference did not stem exclusively from a legitimate regulatory distinction.

In response, the US Department of Commerce issued a new rule in 2013, modifying its requirements for tuna labeled as dolphin-safe under the DPCIA.<sup>8</sup> It is this rule, and the question of whether it brought the US into compliance with Article 2.1, that lies at the heart of the latest round of WTO decisions that we focus on in this article. Mexico requested a review of the modifications made by the 2013 rule, arguing that they did not bring the US into compliance with the provisions of the TBT. The Panel issued a report in April 2015, finding that, even with the modifications made in 2013, parts of the US regulatory regime remained inconsistent with Article 2.1.

The rule kept in place the provision that ETP tuna caught by setting on dolphins could never be eligible for dolphin-safe labeling, but it added requirements for tuna harvested outside the ETP through similar, encirclement methods – so-called purse

4 Dolphin-safe labeling standards, 50 C.F.R. §§ 216.91 and 216.92.

5 *Earth Island Inst. v. Hogarth*, 494 F3d 757 (9th Cir. 2007) (agency decision that purse-seine nets had no significant adverse impact was arbitrary and capricious because, among other reasons, it was based too much on policy rather than science).

6 Panel Report, *United States – Measures Concerning the Importation, Marketing and Sale of Tuna and Tuna Products*, WT/DS381/R, 15 September 2011, para. 7.620 (*US–Tuna II (Mexico)*). The original Panel also determined the US regulation did not offend Article 2.4 of the TBT and declined to reach a decision on GATT claims made by Mexico.

7 Appellate Body Report, *United States – Measures Concerning the Importation, Marketing and Sale of Tuna and Tuna Products*, WT/DS381/AB/R, 16 May 2012 (*US–Tuna II (Mexico)*). The AB affirmed the Panel’s decision on Article 2.4 and made other rulings undergirding its conclusions, such as affirming the Panel’s treatment of the rules governing a *voluntary* labeling scheme as a technical regulation under TBT.

8 78 Federal Register 40,997–41,004 (9 July 2013).

seine fishing methods. Specifically, to be labeled dolphin-safe, tuna caught elsewhere would need to contain certification that no dolphins were killed or seriously injured, just as for tuna caught through any method within the ETP. The rule, though, also contained new requirements for documentation and verification for tuna from within the ETP, and it continued to require an independent observer on boats engaged in tuna fishing in the ETP.

The WTO compliance Panel found that the additional certification, tracking, and verification requirements imposed on ETP purse seine fisheries violated Article 2.1, but that the provision denying eligibility for tuna caught by setting on dolphins did not.<sup>9</sup> In November 2015, the AB reversed the compliance Panel's decision. It found that, by making the absence of setting on a precondition for labeling tuna as dolphin-safe, the US regulations were indeed inconsistent with Article 2.1. Although the AB also reversed the Panel on the additional certification and tracking requirements for the ETP purse-seine fisheries, it examined the entirety of the amended tuna measure and took notice that, despite making regulatory modifications, the US continues to bar most Mexican tuna from dolphin-safe labeling, while at least conditionally allowing similar products to be labeled dolphin-safe.<sup>10</sup> Significantly, the AB found that the more onerous requirements in the US regulatory regime (particularly those for ETP fisheries) did not align with the levels of risk faced by dolphins in different locales and under different methods for tuna fishing. Notwithstanding the US regulatory amendments that imposed new restrictions on purse seine fishing outside the ETP, the AB concluded that because of a perceived lack of alignment between the risks to dolphins and regulatory burdens imposed on Mexican tuna fishers, the US regulatory regime could not be said to reflect legitimate regulatory distinctions.<sup>11</sup>

Following the AB's 21.5 decision, the US Department of Commerce adopted an interim final rule in March 2016, intended to bring the US regulatory regime into full compliance with WTO obligations.<sup>12</sup> Nothing in the interim rule modifies regulations on ETP purse seine fisheries, but it effectively combines categories of fisheries so that the rules on fishing elsewhere are made more consistent with those for ETP purse seine fisheries. For example, it allows the National Marine Fisheries Service (NMFS) Assistant Administrator to make a determination requiring an on-board observer if another fishery either has 'a regular and significant association between tuna and dolphins' or shows signs of a 'regular and significant

<sup>9</sup> Article 21.5 Panel Report, WT/DS381/RW, 14 April 2015 (Compliance (Art. 21.5)). The Panel also made findings with respect to claims under GATT, which we do not focus as they tend to track the analysis under the TBT.

<sup>10</sup> After an extended consideration of whether it could permissibly consider different parts of the US regulatory regime separately, the AB expressed skepticism at the Panel's approach to segmenting its analysis rather than considering all components of a regulatory regime as a whole.

<sup>11</sup> The AB reversed the Panel's findings with respect to the GATT provisions based on reasoning similar to that it applied in assessing TBT Article 2.1.

<sup>12</sup> 81 Federal Register 15,444–15,449 (23 March 2016).

dolphin mortality or serious injury'.<sup>13</sup> The rule seeks to make clear that any vessel, using any fishing method, that intentionally deploys fishing gear against dolphins will be ineligible for dolphin-safe labeling and that all captains must issue certifications similar to what captains of purse-seine vessels operating within the ETP must make.<sup>14</sup>

## 2. Regulatory distinctions: twisting turns

Even this brief recitation of the disputes over the US tuna fish labeling regulation should suggest a considerable degree of complexity to both the underlying regulatory standards and their analysis under applicable trade law. Although we are most interested here in the latest round of the WTO dispute settlement process evaluating the US response to the original AB ruling, we consider the broader treatment of distinctions in the US regulatory regime over the course of the *US–Tuna II* dispute to show how different WTO tribunals have responded to different regulatory distinctions. Our basic argument is that the regulatory regime underlying this dispute affords ample ammunition for a complainant to argue, and a tribunal to find, a lack of ‘even-handedness’ because of the plentiful distinctions embedded in the regulatory regime. In this dispute, several tribunals focused on different distinctions at different stages in the dispute settlement process and found various distinctions insufficiently calibrated to the underlying risk to dolphins.

We accept, at least for our discussion here, the basic conclusion that the US acted in a manner that treated the Mexican fishing industry less favorably – as well as the legal predicates underlying that conclusion. Others have elsewhere questioned some of these predicates, such as whether the US measure was a technical regulation or whether the tuna harvested by Mexico constituted a like product to other tuna (e.g., Mavroidis, 2013). Still other predicates surely merit additional consideration too, such as the compliance Appellate Body’s decision to reexamine the entirety of the US measure again, rather than just asking whether the changes in US law had adequately responded to the original AB decision (Howse, 2015). For our present purposes, we accept the AB’s basic analysis of adverse differential impact, but focus instead on the second step of the AB’s analysis: whether the regulatory distinctions in US law might be justified notwithstanding their differential impact on Mexican fishing operations. This second step was previously articulated by the AB in *US–Clove Cigarettes*, namely whether ‘the detrimental impact on

<sup>13</sup> Ibid.

<sup>14</sup> Under the new rule, ‘captains of all vessels in [ETP large purse seine] fisheries [and large-scale drift-net fisheries] must certify that, no purse seine net or other fishing gear was intentionally deployed on or used to encircle dolphins during the fishing trip in which the tuna were caught, and that no dolphins were killed or seriously injured in the sets or other gear deployments in which the tuna were caught. This revision makes clear that tuna does not meet the dolphin-safe standard if it is harvested by vessels that intentionally deploy fishing gear (regardless of the type) on dolphins’. 81 Federal Register 15,444, 15,446.

imports stems exclusively from a legitimate regulatory distinction rather than reflects discrimination against the group of imported products'.<sup>15</sup>

In *US–Tuna II*, the original AB analyzed the legitimacy of regulatory distinctions based on whether they were well-calibrated with the underlying risks to dolphins, finding that that ‘the United States has not demonstrated that the difference in labeling conditions ... is “calibrated” to the risks to dolphins arising from different fishing methods in different areas of the ocean’.<sup>16</sup> As a result, the AB could not find that ‘the detrimental impact ... on Mexican tuna products stems exclusively from a legitimate regulatory distinction’ nor that the US regime was ‘even-handed in the relevant respects’.<sup>17</sup>

In the latest 21.5 proceeding, the AB again accepted the importance of the same showing, stating that ‘a panel must carefully scrutinize whether the technical regulation at issue is even-handed in its design, architecture, revealing structure, operation, and application in the light of the particular circumstances of the case’.<sup>18</sup> If the measure ‘is designed or applied in a manner that constitutes a means of arbitrary or unjustifiable discrimination, that distinction cannot be considered “legitimate”’.<sup>19</sup>

Throughout the *US–Tuna II* process, the benchmark for the legitimacy or even-handedness of the regulatory distinctions in the US regime rested with *risk*. As the compliance AB observed:

[B]oth the United States and Mexico advanced arguments relating to the respective risks to dolphins associated with different methods of fishing inside and outside the ETP. Mexico sought to establish that tuna fishing methods other than setting on dolphins have substantial adverse effects and that dolphins face risks of mortality or serious injury from tuna fishing outside the ETP that are equal to or greater than those posed to dolphins by fishing within the ETP. For its part, the United States contended that the changes incorporated into the amended tuna measure through the 2013 Final Rule responded directly to the lack of calibration that the Appellate Body found to be responsible for the lack of even-handedness of the original tuna measure. In making its arguments, the United States relied on the findings from the original proceedings, as well as additional evidence, to establish that setting on dolphins in the ETP is ‘particularly harmful’ to dolphins, and that the risks associated with such a tuna fishing method are greater than those associated with fishing methods that are not disqualified from access to the dolphin-safe label under the amended tuna measure.<sup>20</sup>

<sup>15</sup> *United States – Measures Affecting the Production and Sale of Clove Cigarettes*, AB-2012-1 (2012), para. 215. For recent discussion of the origin, meaning, and some consequences of the legitimate regulatory distinction test introduced by the AB, see Houston-McMillan (2016).

<sup>16</sup> Appellate Body Report, *US–Clove Cigarettes*, para. 297.

<sup>17</sup> *Ibid.*

<sup>18</sup> Appellate Body Report, *Compliance* (Art. 21.5), para. 7.31.

<sup>19</sup> *Ibid.*

<sup>20</sup> *Compliance* (Art. 21.5), Appellate Body Report, para. 7.112.

With this foundation, we turn to what distinctions exist within the US tuna labeling regulatory regime which might plausibly need to be shown to be legitimate, before turning to the twisting turns taken by the WTO tribunals in evaluating these distinctions.

## 2.1 *Distinctions in tuna labeling law*

All regulatory regimes contain distinctions. That is how regulations or rules operate (Schauer, 1991). Accordingly, although the US regime governing dolphin-safe labels is based on a short statute – the Dolphin Protection Consumer Information Act – and its rather short implementing regulations,<sup>21</sup> both are replete with distinctions.

At a global level, the DPCIA distinguishes between lawful and unlawful tuna labeling.<sup>22</sup> The statute and implementing regulations then make further distinctions based on: *tuna fishing methods* and *geographic locations* for tuna fishing; conditions for *eligibility*; and requirements for *certification* for dolphin-safe labeling. These eligibility conditions and certification requirements are themselves contingent on further distinctions based on fishing methods, geographic locations, and even the size of fishing vessels. Let us consider each of these distinctions further so as to understand better how they were assessed by the WTO tribunals.

### 2.1.1 *Tuna fishing methods*

The *US–Tuna II* dispute revolves largely around purse seine fishing, a technique which involves a boat dropping a net vertically into the water and then encircling a school of tuna with it, basically wrapping a ‘bubble’ of net around the school, and then picking up the net with the tuna in it. Other species of fish and marine wildlife will typically be caught in the nets too – so-called bycatch. Dolphin bycatch can be more likely to occur in certain parts of the ocean, particularly the ETP, because tuna for some unknown reason especially like to swarm around dolphins there.

With purse seine fishing, fishers can determine where to ‘set’ or encircle their nets in different ways. Since tuna like to swim around dolphins in the ETP, an effective way of finding large tuna there can be to lay or set the net intentionally around a group of dolphins, which makes the tuna easier to spot and ensures a robust catch of tuna. However, this setting on dolphins also leads to death, injuries, and stress to dolphins as they become entangled in the net. Other ways of setting avoid dolphins, as they rely on logs or ‘fish aggregation devices’ to attract tuna – or they just involve searching harder for schools of tuna without dolphins in their midst, as must be done in other parts of the ocean (Table 1).

In addition to purse seine fishing, tuna can be harvested through driftnets, which are dragged horizontally through the water and pulled by a boat. Driftnets, though,

<sup>21</sup> The tribunals also included *Hogarth*, *supra* note 5, as part of the measure.

<sup>22</sup> 16 USC. 1385(d).

Table 1. Methods of tuna fishing

---



---

1. Purse seine (§1385(d)(1)(B)&(C))
• Dolphin setting ('Setting on')
• Log or FAD (Fish Aggregation Device) setting
• School setting
2. Driftnet (§1385(d)(1)(A))
3. Other methods (§1385(d)(1)(D))
• Long line
• Trolling
• Pole & line

---



---

entangle and capture a large volume of bycatch, including dolphins. A variety of other assorted methods involve dragging or trolling single lines; these, too, can yield bycatch, sometimes of dolphins too, but generally driftnet fishing and purse seine fishing by setting on are the ones that result in the most dolphin bycatch.

### 2.1.2 *Geographic locations*

The US regulatory regime draws distinctions based on fishing locations: the Eastern Tropical Pacific Ocean (ETP); areas outside the ETP; the high seas; and 'other fisheries'. The ETP is defined by the DPCIA to extend westward to 160 degrees longitude from parts of North and South America. This is significant for the *US–Tuna II (Mexico)* dispute for two reasons. First, although scientists do not really know why, dolphins and tuna show a much greater propensity to swim together in the ETP than they do elsewhere in the ocean. Second, Mexican fishing operations have long used purse seining fishing methods in the ETP. Parts of the ETP lie off some of Mexico's western coastline; however, nine other countries' coastlines abut the ETP and the vast majority of it lies in international waters, where other countries' boats can fish.

### 2.1.3 *Certifications*

If tuna is to be labeled as dolphin-safe, US regulations call for one or more certifications, depending on where tuna is harvested and how. One of these certifications is that dolphins were not intentionally targeted – i.e., no setting on dolphins. The second is a certification that no dolphins were killed or seriously injured in a haul of tuna. For tuna harvested within the ETP through purse seine methods, both certifications must be made. At the time of the original AB decision, purse seine fishing boats *outside* the ETP only needed certification of no setting on.

### 2.1.4 *Certifiers*

Depending on where and how tuna is harvested, those responsible for filing the necessary certification must be the captain of the vessel, an observer (such as

someone who is part of an international dolphin protection program), or both. Within the ETP, both are required. At the time of the original AB decision, purse seine fishing boats outside the ETP needed only certification by the captain.

### 2.1.5 Findings

US law distinguishes between various kinds of findings the Secretary of Commerce can make about particular fisheries, with regulatory implications hinging on whether a finding is made or not. For example, if the Secretary makes a finding that dolphins in the ETP no longer suffer a ‘significant adverse impact’ from tuna fishing, then purse seine fishing boats would no longer need to certify that there was no setting on to be able to label their tuna as dolphin safe.<sup>23</sup> If the Secretary finds a ‘regular and significant association’ between dolphins and tuna swimming together in an area outside the ETP, then purse seine vessels need to certify both ‘no setting on’ and ‘no killing or serious injury’. The same is true for non-purse seine methods in another location, provided the Secretary makes a finding of no ‘regular and significant mortality’ of dolphins in that location. It should be noted that no such authoritative findings yet have been made – so the current structure merely provides for these findings as possibilities.

### 2.1.6 Size of purse seine fishing vessels

The DPCIA places labeling conditions on ‘a vessel using a purse seine net’ within the ETP.<sup>24</sup> By contrast, US implementing regulations have, from the earliest days, only applied the ETP purse seine conditions to vessels with a carrying capacity larger than 400 short tons (363 metric tons). The original Panel acknowledged that no conditions whatsoever applied within the ETP to smaller vessels.<sup>25</sup> The way the regulations are written, the lack of any conditions placed on small boats means that any tuna they catch, by any methods, even if dolphins are killed in the process, can be labeled as dolphin-safe without offending the DPCIA. The original AB decision made no mention of this distinction. The compliance Panel did pick up on it briefly, implying that the US regulatory change in 2013 had closed up this proverbial loophole. But, in fact, the regulatory change made by the US did not alter any language related to the size of the ETP purse seine vessel. Upon questioning by the Panel, the US did offer an *understanding* that smaller vessels would be treated in a manner parallel to purse seine vessels in non-ETP waters, an understanding accepted by both the Panel and the AB but not grounded in any regulatory language.

<sup>23</sup> The Secretary had made such a finding prior to the initiation of this dispute settlement proceeding, but the Ninth Circuit had set it aside as arbitrary and capricious under the Administrative Procedure Act. *Hogarth, supra* note 5.

<sup>24</sup> 16 USC. 1385(d)(1)(C).

<sup>25</sup> Panel Report, *US–Tuna II (Mexico)*, table at pp. 7–8.

## 2.2 *Twisting turns in US–Tuna II*

This overview of the US–dolphin-safe regulatory regime reveals six potential distinctions – and sub-distinctions within them – that could potentially be scrutinized for legitimacy or even-handedness. Some of these distinctions – but by no means all – have figured into the WTO tribunals’ analyses. In fact, at each major step along the dispute settlement process, the central distinction of concern to the tribunal has shifted, with new distinctions of concern emerging along the way. The treatment of regulatory distinctions throughout the *US–Tuna II* proceedings appears to have been ad hoc, if not opportunistic. This may be a sign, of course, of successful lawyering on behalf of claimants seeking to take another bite at the apple each time; however, it also evinces an apparent lack of a firm, principled foundation upon which to determine what distinctions matter and what needs to be shown to demonstrate that the relevant distinctions are sufficiently calibrated and legitimate.

Although the original Panel quite explicitly focused only on the distinction between purse seine fishing *with* and *without* dolphin-setting, the original AB went in a new direction by focusing on the difference in the required certification that dolphins were not killed or seriously injured between *ETP* and *non-ETP* purse seine fishing. After the US removed the latter distinction through a regulatory change, the Article 21.5 AB shifted yet again to another distinction, this time in who must certify: *captains and observers* or *captains only*, as between *ETP* versus *non-ETP* purse seine fishing, respectively. As Howse (2015) has suggested, ‘it is truly difficult to fathom the AB’s reasoning’ in its 21.5 decision. Let us consider the twisting path that led to this latest decision.

### 2.2.1 *Original Panel*

The original Panel recognized that ‘the US dolphin-safe labeling provisions involve a number of regulatory distinctions’.<sup>26</sup> But it focused on ‘[t]he regulatory distinctions at issue’ as those involving ‘setting on dolphins as opposed to not setting on dolphins’.<sup>27</sup> This limited focus on ‘setting on’ versus ‘not-setting on’ was surely a function of what the parties had argued, which in turn was no doubt affected by the fact that ‘at least two thirds of Mexico’s purse seine tuna fleet fishes in the ETP by setting on’.<sup>28</sup> By contrast, most US fishing operators no longer use setting on techniques.<sup>29</sup> The Panel concluded that the measures did not violate Article 2.1 because anyone could choose to adopt a method of tuna fishing that did not involve setting on dolphins. (In fact, not only have US fishing operators abandoned the practice in the ETP, but also an influx of fishing boats from Ecuador has entered the ETP in the years following the DPCIA, which do

<sup>26</sup> Panel Report, *US–Tuna II (Mexico)*, para. 7.304.

<sup>27</sup> *Ibid.*, p. 186.

<sup>28</sup> *Ibid.*, para. 7.314.

<sup>29</sup> *Ibid.*, para. 7.316.

not use the technique.) The Panel did acknowledge that ‘the measures at issue involve further distinctions, in particular in the certification requirements for different areas, depending on the status of the fishery, the type of vessel[,] and fishing method used’.<sup>30</sup> However, the Panel did not consider any of those other distinctions for the purpose of its Article 2.1 analysis: ‘Mexico has indicated that these differences in treatment are not the factual basis for its discrimination claims. We therefore do not address them in this context.’<sup>31</sup> Some of those other distinctions were addressed by the Panel in another context, namely its treatment of Article 2.2, where it did find the US had acted inconsistently with WTO obligations.

### 2.2.2 *Original AB*

When the AB undertook its review of the Panel’s decision, it flipped ultimate conclusions, finding the US violated Article 2.1 but not Article 2.2. Significantly, setting versus not-setting was no longer the only distinction Mexico advanced in support of its claim under Article 2.1. Rather, Mexico emphasized that its fishing operations within the ETP, in order to comply with the dolphin protection program under the Agreement on the International Dolphin Conservation Program (AIDCP), already had on-board observers to certify that dolphins were not killed or seriously injured, something also required by US law (in addition to no setting on) for purse seine fishing in the ETP. By contrast, similar purse seine fishing operations outside the ETP did not need to have anyone certify that no dolphins were killed or seriously injured in order to qualify for the US–dolphin-safe label, only that no setting on occurred.

The AB relied on much of the Panel’s Article 2.2 findings to bolster its conclusion that the US had violated Article 2.1. The AB concluded that the US regulatory regime discriminated against Mexico, and when it turned to consider the possibility of a legitimate regulatory distinction, the distinction was no longer setting on versus no setting on, but instead certification versus no certification of an absence of dolphin kills and serious dolphin injuries. The AB asked ‘whether the detrimental impact of the measure stems exclusively from such a [legitimate] distinction rather than reflecting discrimination’.<sup>32</sup>

The AB pointed to uncontested Panel findings that dolphin killings do occur with fishing outside the ETP, whether through purse seine fishing or other fishing methods. The Panel had found that the volume of dolphin bycatch outside the ETP was ‘significant’, while the US argued that the level outside the ETP was vastly lower than within the ETP, claiming that the Panel mistakenly lumped together studies of non-ETP driftnet fishing with those of purse seine fishing and other techniques. (Since driftnet fishing is categorically precluded from the

30 Ibid., para. 7.309.

31 Ibid.

32 Appellate Body Report, *US–Tuna II (Mexico)*, para. 284.

dolphin-safe label, the US argued that those data are irrelevant to the dispute with Mexico.) The AB rejected the US's reasonable concerns and relied on the Panel's finding (again, under its Article 2.2 analysis) that the risks to dolphins from fishing outside the ETP are 'at least equivalent' to those dolphins face from purse seine fishing inside the ETP. These findings told the AB that the US regulatory regime was not 'even-handed' or 'calibrated'.

### 2.2.3 *Article 21.5 AB*

After the US amended its regulations to state that tuna harvested in any fishery outside the ETP would be required to be accompanied by a 'no killing' certification, the compliance AB agreed with the compliance Panel that, in making this change, the US had eliminated the problem with that particular distinction. If certification were the only issue, the AB recognized it would have needed to find that the US had satisfied the even-handedness test and would be in compliance with Article 2.1.<sup>33</sup> However, the AB determined that, in its Article 21.5 review, it should reconsider the entire regulatory profile with this change. In particular, the test should be whether the 'differences in the relative risks of harm to dolphins from different fishing techniques in different areas of the oceans explain or justify the differences in the certification requirements and the tracking and verification requirements applied inside and outside the ETP large purse-seine fishery'.<sup>34</sup> No longer was the distinction just one between setting on dolphins versus no setting on, nor between a 'no killing' certification in the ETP but not outside it. Now the US needed to find differences in 'relative risks' from the different fishing methods and different locations, and then use that risk assessment to justify all the differences in its regulatory regime. Howse (2015) raises a valid question as to why this comprehensive risk accounting had not been viewed as the test before. All that the Article 21.5 AB could muster, apparently, was to say that 'the question as to the relative risk profiles ... has become more acute'.<sup>35</sup> Despite the apparent increase in acuteness, the AB found that the totality of the risk evidence presented before the Panel was either contested or incomplete. Consequently, the Panel had been unable to 'gauge properly the overall relative risks or levels of harm to dolphins ... in order to assess whether the differences in the dolphin-safe labeling conditions under the amended tuna measure are appropriately tailored to, and commensurate with, those respective risks'.<sup>36</sup>

However, that absence of risk evidence did not stop the AB from proceeding to rely on risk and regulatory calibration to find the US still to be out of compliance. Despite its emphasis on the need for relative risk profiles, the AB satisfied itself that

33 Appellate Body Report, Compliance (Art. 21.5), paras. 7.241–7.242.

34 *Ibid.*, para. 7.229.

35 Appellate Body Report, Compliance (Art. 21.5), para. 7.251.

36 *Ibid.*, para. 7.252.

it could proceed on the basis of ‘other features’ of the regulatory regime ‘that are not dependent on an assessment of the relative risks associated with different fishing methods in different areas of the oceans’.<sup>37</sup> It called attention to a difference in the wording of the US law about possible findings the Secretary of Commerce could make to require that both a captain and an observer, and not just a captain, certify to ‘no setting on’ and ‘no killing’. For non-ETP purse seine fishing, the finding that would trigger the observer requirement would need to be one of ‘regular and significant *association*’ of dolphins and tuna, while, for non-ETP non-purse seine fishing, the trigger would be ‘regular and significant *mortality*’. The AB rejected the US position that mortality follows directly from association with purse seine fishing. The AB wanted to see a ‘comparable regulation of a risk scenario’ in which mortality was found directly in both settings – not using a proxy (association) in one but not the other.<sup>38</sup> As a result, the US regulations ‘do not provide for the substantive conditions of access to the dolphin-safe label ... in all circumstances of comparably high risks’.<sup>39</sup> The AB concluded ‘it has not been demonstrated that the differences in the dolphin-safe labeling conditions under the amended tuna measure are calibrated to, or commensurate with, the risks to dolphins arising from different fishing methods in different areas of the oceans’.<sup>40</sup> Despite language about ‘comparably high risks’, the Article 21.5 Appellate Body’s reasoning was all based on logical categories without any empirical risk assessment data, which the AB acknowledged the Panel had failed to obtain.

### 2.3 Roads not taken

Just as magicians pull rabbits out of hats, over the course of the *US–Tuna II* proceedings, adjudicators kept pulling new distinctions out of the US regulatory regime. It was almost as if no good distinction could go to waste. Yet, despite what might look like a certain kind of eagerness to find new regulatory distinctions around each bend, there are at least two obvious distinctions that never emerged as instrumental in the dispute. These roads not taken in the WTO proceeding can reveal something important about the challenge of principled adjudication of even-handedness. Just as concerns about the WTO process arise from the twisting turns in the distinctions that *did* become central to the tribunals’ Article 2.1 analysis, distinctions that *were not* examined also reveal a degree of incoherence to the WTO’s invocations and analyses of regulatory distinctions. Why some distinctions and not others?

37 Ibid., para. 7.254.

38 Ibid., para. 7.259.

39 Ibid., para. 7.266.

40 Ibid.

### 2.3.1 Driftnet fishing

Especially if risk is supposed to be the lodestar, it is surprising that driftnet fishing did figure prominently anywhere in the dispute. The original AB did refer, in passing, to how the DPCIA says that the dolphin-safe label is not available for tuna caught ‘on the high seas by a vessel engaged in driftnet fishing’.<sup>41</sup> But there is no corresponding provision in the statute for driftnet fishing *not* on the high seas. As the original AB stated, ‘[w]e note that while the measure at issue stipulates that tuna caught using driftnets on the high seas is not eligible for a “dolphin-safe” label, it grants access to the label to tuna products containing tuna caught with driftnets in exclusive economic zones’.<sup>42</sup> Mavroidis (2013) has perceptively noticed this fleeting reference in the original AB decision, pointing out that ‘it allowed tuna caught with driftnets in its EEZ (exclusive economic zone) to be called “dolphin-safe” although most serious threats for dolphins occur in coastal areas, i.e., within EEZ’.

Such a regulatory distinction for driftnet fishing – prohibited on the high seas, while permitted in the EEZ – would seem completely unjustifiable on the basis of risk, given both how destructive driftnet fishing is to marine mammals in general and within the EEZ in particular. One would reasonably think that this disparity should surely have proven devastating to the US case. And yet, perhaps because the heart of the *US–Tuna II (Mexico)* dispute centered on purse seine fishing, it received only fleeting attention – attention given only in the course of dismissing US claims that the original Panel erred in referring to studies that included data on driftnet fishing. Perhaps because it was just a passing reference, the AB did not recognize that, even though the US *statute* limits the driftnet prohibition to the high seas, the corresponding *regulations* implementing the statute make no such limitation. Even at the time of the AB’s decision, the regulations clearly applied simply to any ‘vessel engaged in large-scale driftnet fishing’ – period.<sup>43</sup>

The point here is mainly to indicate another instance of the seeming ad hoc nature of the treatment of risk and regulatory distinctions, not to dwell on the AB’s error in characterizing US law on driftnet fishing. The point is that this was a clear road not taken. What is surprising is that, given the original AB’s (erroneous) understanding of US law as well as both ABs’ emphasis on consistency in the treatment of risk, none of the panels ever focused on this distinction.

41 16 USC. § 1385(d)(1)(A).

42 Appellate Body Report, *US–Tuna II (Mexico)*, para. 270.

43 50 C.F.R. §216.91(a)(3). This language was in effect at the time of the AB’s decision, and even in the most recent reorganization of the CFR provisions effectuated by the March 2016 interim rule, this language remains unchanged. Driftnet fishing simply does not qualify for dolphin-safe labeling, no matter where it occurs.

### 2.3.2 *Dolphin species*

Given the emphasis on risk, it was also striking never to see distinctions drawn between different species of dolphins, their habitats, and population health. Granted, the parties themselves appear not to have made much of species' distinctions. But if the US labeling system were to be truly calibrated to the risks that fishing poses to dolphin populations, the statute or the regulations presumably would need to take into account differences across species. Yet virtually nothing in the US labeling regime makes any distinctions between species, their habitats, or population health. The statute does call for the Secretary of Commerce to study the 'depleted dolphin stock' in the ETP, but other than that single reference – which itself specifies nothing about species – the legislation says nothing about particular dolphin stocks. Furthermore, the implementing regulations do not mention differences in species either, even though other regulatory provisions in the same part of the US Code of Federal Regulations list specific species of dolphins as 'depleted' and meriting special protection under other regulatory provisions.<sup>44</sup> The very existence of mention elsewhere in US law of the biological reality that different types of dolphins face different types of population risks makes plain a second road not taken by the tribunals: calling out the US *failure* to make legal distinctions about species risks. It is hard to see how a regime could ever be said to be well calibrated to dolphin risks without recognizing biological realities.<sup>45</sup>

## 3. Challenges in calibration

The WTO tribunals' twisting turns and roads not taken are clearly visible from a risk-based perspective on regulation. Yet, strikingly, risk-based rhetoric itself pervaded the various panel decisions throughout the *US–Tuna II* dispute, undoubtedly stemming from several sources. The US, for instance, advanced risk arguments in defense of its regime. But the pull of a risk logic in this dispute also likely stemmed from a deep interest in performance-based regulation that infuses the field of trade and regulation more generally. It is surely no accident that Article 2.8 of the TBT, while not invoked in the tuna dispute, urges WTO members to draft technical regulations 'in terms of performance' whenever possible. Representatives to the WTO's TBT Committee 'have stressed ... the advantages of performance-based regulation'<sup>46</sup> and urged that standards as well as regulations

44 E.g., 50 C.F.R. § 216.15.

45 The statute does, of course, give the Secretary of Commerce considerable implementing discretion, so it is conceivable that in practice the Secretary could put in place some more fine-grained requirements. But even a Secretary sensitive to and concerned about variation in species' health would undoubtedly find the basic structures of the regime – conditions, certifications, and verification – all invariant with respect to dolphin species.

46 WTO Committee on Technical Barriers to Trade, *Compilation of Sources on Good Regulatory Practice*, G/TBT/W/341, 13 September 2011, p. 5.

‘should be performance based rather than based on design or descriptive characteristics’.<sup>47</sup> Throughout the world, we can see considerable interest in stating regulatory commands in terms of outcomes (Coglianese, 2016; Hutter, 2016).

Performance-based regulation’s close cousin is what has been widely referred to as risk-based regulation, which, as its name implies, calls for regulatory decisions to be based on careful consideration of risk – formally defined as a probability multiplied by the severity of a problematic consequence (Paoli and Wiles, 2015). Because any public policy problem can be redefined in terms of the risk of that problem – that is, either its probability of occurring or its severity – risk-based regulation can be, and is, viewed as attractive for financial regulation, transportation safety, homeland security, environmental protection, and any number of other regulatory domains. Part of risk-based regulation’s appeal stems from its promise to provide more objective and coherent regulatory decisions. The great advances in risk assessment and analysis in recent decades have, not unreasonably, given rise to expectations for improved rationality in regulatory decisions.

Rationality is what WTO dispute settlement bodies purport to assess when trying to determine whether an otherwise seemingly discriminatory regulation is nevertheless based on a sound, objective rationale – as opposed to something ‘arbitrary’. Rationality can be understood as an alignment between means and ends, a conception that helps explain the appeal of *performance*-based regulation. After all, if a challenged regulation is stated in terms of its ends, rather than means, then it would be impossible to conclude that the regulation could be arbitrary or unjustified under the TBT Agreement, as long as the ends are permissible ones, such as ‘protection of human, animal or plant life’ as stated in its preamble.

Affinity for performance-based regulation naturally creates an affinity for risk-based regulation. The mandatory ends that a performance-based regulation imposes can obviously be stated in terms of risk – or in other outputs related to risk, such as, say, emissions of pollution. However, even if a regulation is not defined in terms of a risk-related output or outcome, but instead mandates a specific means or action, risk-based reasoning can guide the selection of means that are well-aligned with the underlying risks aiming to be reduced or managed. Mandated means, in other words, ought to be ‘calibrated’ with desired ends.

When it comes to protecting dolphins, some regulatory theorists might posit a simpler tuna labeling regime with distinctions based directly on dolphin risks. In principle, a regulation could simply state that, no matter the means of fishing or its location, no dolphins can be killed or injured in the harvesting of any tuna that is to be labeled as dolphin-safe. After all, to have a label state that tuna is ‘dolphin-safe’ is to make a claim about a particular *outcome* of harvesting:

<sup>47</sup> WTO Committee on Technical Barriers to Trade, *Decisions and Recommendations Adopted by the WTO Committee on Technical Barriers to Trade since 1 January 1995*, G/TBT/1/Rev. 12, 21 January 2015, p. 49.

namely, safe, unharmed dolphins.<sup>48</sup> Such a no-harm labeling rule would offer the flexibility of performance-based regulation, as it would not say anything about *how* tuna can or should be caught, just that whatever means a fishing crew uses, it would need to maintain the outcome of dolphin safety.

Would such an approach be risk-based as well? In one sense, obviously so. It would condition labeling on the minimization of any risk to dolphins. But in another important sense, such a simple law would obscure an important feature of risk in this context: population risk versus individual risk. Consider, for sake of analysis, if the US had adopted a flat no-harm ban – not a labeling law, but a ban on harming any dolphin. Such a ban would maximally protect individual dolphins, but it would not necessarily be well-calibrated with respect to dolphins' population risk. A flat ban on harm to any individual dolphin would be too strict in some parts of the ocean where dolphins are thriving and the species overall could withstand harm to some individual dolphins. It could demand too much protection (and too high a regulatory burden) when not needed from a population standpoint.<sup>49</sup>

Of course, sometimes laws do impose flat bans. The US Endangered Species Act bans any harming, taking, and harassing of individual members of species that have been determined to be threatened or endangered. But such a law affords this individual protection only upon a finding that a species is threatened. Absent a condition of overall species endangerment, a flat ban on harm to individual members of a species would presumably be out of alignment with population risk. The ideal, perfectly calibrated performance-based regulatory standard to manage *population* risk would never take the form of a ban on any and all harm to individual dolphins. It would be a rule written in performance terms with an outcome stated in population-risk terms.

As a practical and analytical matter, of course, such an approach would be infeasible, if not impossible, to carry out. Harm to an individual dolphin can be observed; the risk of harm to a dolphin *population* is not directly observable and cannot be embedded in any enforceable legal obligation or condition. In the real world, a dolphin *species* protection law will necessarily be based either on observable *means* of fishing (e.g., mandating use of fish aggregation devices, prohibiting driftnets) or on discrete *outcomes* (that is, bycatch quotas, which in some instances might need to be zero but in other instances could be greater than that). Either approach would presumably need to vary depending on the level of viability of

<sup>48</sup> Of course, a labeling regime does not guarantee safety to dolphins. The label is not mandatory, so tuna can continue to be caught in ways that harm dolphins as long as the producers are able to and willing to sell their tuna without the dolphin-safe label. In the widest sense, the ultimate outcome of a performance-based *labeling* law is a better-informed consumer, not a protected dolphin.

<sup>49</sup> Removing the labeling component of the regulatory regime, a flat ban on harming any dolphins would come at the cost of tuna as a source of food, as presumably any kind of tuna fishing poses some unintentional, residual, even *de minimis*, level of risk of harm to dolphins. To be welfare-maximizing, regulation would not necessarily prevent *all* harm to dolphins, but just reduce risk to an optimal level, where the marginal costs of harm prevention equal the marginal benefits.

fish populations in certain parts of the ocean, as well as vary over time because species' viability would undoubtedly change due to levels of fishing as well as other factors in dolphins' ecosystems. With perfect information, a resource manager could issue individual bycatch quotas that would be precisely estimated and dynamically updated for each location in which a fishing boat might operate. Perhaps with the assistance of machine learning algorithms, each fishing boat could receive a precise bycatch limit every day or every week, a limit precisely targeted for its given location and computed to achieve resource management goals.<sup>50</sup>

Such an approach is not feasible in the world we inhabit today, with its very real and substantial information costs. But in imagining this way of operationalizing population risk in terms of perfectly fine-tuned individual-level bycatch quotas, we can now envision an ideal state that would clearly meet the AB's test of a regime where conditions are 'appropriately tailored to, and commensurate with, those respective risks'.<sup>51</sup> We can also see another, more fundamental problem that arises if calibration is taken to this extreme: it no longer calls for *regulation*. The issuance of bycatch permits to individual vessel operators would be individualized adjudications, not a type of general rule-making. Such case-by-case decisions based on empirical forecasts would be rather the opposite of rules. Rules make general distinctions. They are, by definition, generalizations (Schauer and Zeckhauser, 2007). And they contain some irreducible level of over- or under-inclusiveness, at least in certain locations, for some types of behavior, and at particular time periods. Once the choice is made to create rules, whether as conditions or commands, the distinctions they contain can never be fully and perfectly 'calibrated', for if they were, they would no longer be *regulatory* distinctions.

When it comes to 'calibration' under Article 2.1, then, there exists no Holy Grail of precision. The key question should not so much be, 'Is the regulatory distinction calibrated?' The question should be the prior, and more difficult one, namely, 'How finely calibrated must the regulatory distinction be?' The answer to this question must presumably submit to reason. The important point for our purposes is to recognize that calibration will be, and necessarily must be, to some degree imprecise and uncertain with any rule.

When policymakers and scholars go looking for a risk-based approach to regulation, they must understand what they can realistically expect to find. Too seldom do they appreciate a host of questions that call out for resolution when regulating risk. Rather than providing a clear, principled guide for regulatory decision making, a call for risk-based regulation only necessitates making other choices which hold real policy implications but for which there are often few uncontested, determinative principles. Worst of all, when advocates of risk-based regulation fail

<sup>50</sup> For further discussion of similar, perhaps not-too-fictional, possibilities, see Coglianese and Lehr ([forthcoming](#)).

<sup>51</sup> Appellate Body Report, Compliance (Art. 21.5), para. 7.252.

even to acknowledge these choices, they end up making ad hoc decisions based on unexamined assumptions.

What are some of these choices that must be resolved, even with risk-based regulation? We offer several major examples.

1. *What factors should be taken into account in assessing risks and calibrating them with regulatory distinctions?* Risks are not self-evident. Since risk is a function of both *probability* and the *severity* of harms, these two facets will matter, although sometimes one may matter more than the other. In some cases, the severity of harm may matter greatly. For dolphins, the severity of harm was a point of contention throughout the dispute settlement proceeding. Granted, killing is killing, but dolphins may also be harmed just from the stress of the chase, even if they receive no visible injury. Of course, severity of harm to a *population* as opposed to an individual animal remains a challenging issue to pin down. Even if the absolute number of dolphins killed is high, this could still be small relative to the overall population of dolphins. Finally, *exposure* matters too. As was uncontested in the proceeding, tuna do not associate with dolphins to the same extent in every region of the ocean, so the same type of fishing will not present the same risk exposure to dolphins in different parts of the sea.
2. *What principles of risk management should be applied in a risk-based regulatory regime?* Saying that regulation should be based on or calibrated to risk does not fully explain exactly *how* risk should enter into a justification for regulation. All of the following five principles could be applied in a ‘risk-based’ regulatory regime, but they will not always lead to the same policy outcomes: (a) Prevent the most severe harm; (b) Minimize risk (i.e., probability times severity of harm); (c) Avoid unacceptable risk; (d) Control risk only to the extent feasible; and (e) Maximize net-benefits (i.e., benefits minus costs) (Coglianese, 2015). Does it matter which of these principles a state chooses for the purpose of even-handedness or calibration under Article 2.1? Or is it merely that WTO members or the AB (or both) must articulate some principle and apply it faithfully?
3. *What role should administrative and pragmatic concerns play in justifying regulatory decisions?* Regulators have limited resources, and administrative constraints can legitimately affect the design of a regulatory regime. For example, no regulator can directly oversee every fishing vessel. In addition, there are costs both to the government and the private sector in imposing observer conditions and other verification and tracking requirements. Should these matter? Furthermore, regulators may be cognizant of other social or business imperatives that are likely to affect the behavior of those they oversee: e.g., industry norms, background liability, consumer pressures. These administrative and pragmatic considerations cannot be placed into units of risk, but ultimately they may matter considerably in the design of a regulatory regime (Coglianese, 2002).
4. *What to do when uncertainty exists, as it almost always does?* Uncertainty is different from risk. Risk has known probabilities, whereas under uncertainty either the probabilities or outcomes (or both) are not fully known. The compliance AB

found itself faced with uncertainty, stymied by the contested and incomplete evidence about the relative risks to dolphins from different types of fishing in different parts of the ocean. But this was hardly an exceptional predicament. Uncertainty exists in the real world of policy decision-making all the time. To be sure, analytic techniques exist to estimate levels of uncertainty, i.e., the confidence we can have in risk estimates (Finkel, 1990). But uncertainty cannot always be measured consistently or clearly; sometimes regulators are even uncertain about their uncertainty. What should a WTO body expect in terms of coherence with respect decision-making under uncertainty? Does uncertainty mean that a defendant state is justified in regulating just in case (as under the precautionary principle)? The US argued, after all, that uncertainty over how well the dolphin population was recovering in the ETP justified its continued distinct treatment of that region. How can a WTO tribunal determine when a defendant state's reliance on uncertainty is legitimate?

5. *Who bears the burden of demonstrating risk coherence or incoherence?* Given that an inquiry into calibration arises after a determination is made of less favorable treatment, calibration is set up as a defense, at least according to the AB's analysis. The defendant, then, bears the burden of proof. But there are broader, related questions. Can a democratic choice by a sovereign state suffice to meet that burden and provide an acceptable reason for choices about risks? Is a regulatory distinction legitimate if it is based on a public's fears, regardless of whether they are grounded in science? And which institution has the better knowledge about and ability to assess risks anyway: a nation's regulator or WTO panels? Since an inquiry into regulatory distinctions arises after a finding of less favorable treatment, does it make better sense to give little to no deference to the defendant state and its regulators?

This last question raises an underlying conundrum about risk and regulatory calibration at the second step of Article 2.1 analysis. The AB requires that differential treatment must stem 'exclusively' from a legitimate regulatory decision. But exclusivity raises the bar very high, imposing a potentially impossible practical if not legal burden of proof on the defendant. It will be hard to demonstrate exclusivity, for several reasons. First, if a trade differential is working to the disadvantage of the plaintiff, this will always cast suspicion on what the defendant has done. Second, the range of risk principles and factors noted above means that domestic regulators may often (perhaps always) be able to couch their distinctions in the language of risk and claim that they are calibrated to risks – even though this does not mean that those factors are *exclusively* the reason why the distinctions matter. In real-world policymaking, there can exist, and probably always will exist, a confluence of public interested or even-handed rationales *and* the advancement of private interests or trade discrimination.<sup>52</sup>

<sup>52</sup> Some economists call this a 'Baptist-bootlegger' connection (Smith and Yandle, 2014).

Issues like these not only make principled regulation difficult, they make principled *risk*-based regulation difficult too. The range of principles, choices, and factors that underlie even a risk-based approach to regulation also gives reviewing bodies a considerable number of opportunities to raise factors that either were not articulated well or never nailed down at all. Moreover, what looks coherent when some principles or factors are invoked can be made to look incoherent by another framing (Coglianese, 2002). Absent any fixed, clear guidance, regulatory distinctions might be predictably made to look incoherent.

#### 4. Conclusion

The precision implied by terms like ‘calibration’ or ‘risk-based’ regulation can be deceptive. Indeed, as suggested in Part II of this article, if the aim of pursuing a risk-based approach to regulatory review is to achieve a level of principled coherence in legal decision-making, the experience with the *US–Tuna II* proceedings shows that there are no guarantees against incoherence or ad hoc decision-making. Rather than a risk-based approach constraining WTO decision-making, it might even reasonably appear that an emphasis on risk unshackled adjudicators from principled constraints. As soon as one distinction was addressed, there seemed to be another problematic one to be found. The roads not taken in the proceeding reinforce the conclusion that a risk-based approach to decision-making hardly eliminates discretion.

After the compliance proceeding ran its course, the US government yet again revised its regulatory regime in an effort to adhere to its WTO obligations. It has done so by attempting to blend the incompatible language that the Article 21.5 AB found objectionable, namely the differences between ‘regular and significant association’ and ‘regular and significant mortality’. The US regulatory position would now appear consistent, having addressed the concerns expressed by both AB decisions. And yet, the question remains: Will what the US has currently done be enough? That question, we suggest, cannot be definitely answered in the affirmative. Our review of the *US–Tuna II* dispute leads us to speculate that, if further adjudication continues, another Panel might very well return to the issue and begin again to focus on another set of distinctions or, perhaps, to press that the US has still not done enough to eliminate the uncertainty over whether its overall regulatory scheme is truly risk-based.<sup>53</sup>

In identifying the potential for risk-based rhetoric to disguise ad hoc decision-making, and in raising questions that must be addressed to make risk-based decision-making more coherent, we should not be taken to imply that the WTO dispute settlement process cannot and should not be improved in dealing with cases like *US–Tuna II*. On the contrary, calibration, risk analysis, and benefit-

<sup>53</sup> By June, 2016, the Dispute Settlement Body had referred the dispute back to a panel for further Article 21.5 review.

cost analysis all have the potential to play useful roles both in domestic regulation and in TBT Article 2.1 cases. But using these concepts and tools to make legal and policy decisions less arbitrary will require at a minimum that the Appellate Body provide more careful and coherent guidance to WTO members and to panels than it has done so far.

## References

- Beck, U. (1992), *Risk Society: Toward a New Modernity*, London: Sage Publications.
- Boyd, W. (2012), 'Genealogies of Risk: Searching for Safety, 1930s–1970s', *Ecology Law Quarterly*, 39: 895–987.
- Coglianesi, C. (2002), 'Bounded Evaluation: Cognition, Incoherence, and Regulatory Policy', *Stanford Law Review*, 54: 1217–1238.
- (2015), *Listening, Learning, and Leading: A Framework of Regulatory Excellence*, <https://www.law.upenn.edu/live/files/4946-pprfinalconvenersreportpdf>.
- (2016), 'Performance-Based Regulation: Concepts and Challenges', in F. Bignami and D. Zaring (eds.), *Comparative Law and Regulation: Understanding the Global Regulatory Process*, Cheltenham: Edward Elgar.
- Coglianesi, C. and D. Lehr (forthcoming), 'Regulating by Robot: Administrative Decision Making in the Machine Learning Era', *Georgetown Law Journal*.
- Finkel, A. M. (1990), *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers*, Washington, DC: Resources for the Future.
- Houston-McMillan, J. (2016), 'The Legitimate Regulatory Distinction Test: Incomplete and Inadequate for the Particular Purposes of the TBT Agreement', *World Trade Review*, 15(4): 543–562.
- Howse, R. (2000), 'Democracy, Science, and Free Trade: Risk Regulation on Trial at the World Trade Organization', *Michigan Law Review*, 98: 2329–2357.
- (2015), 'The Tuna/Dolphin Appellate Body 21.5 Ruling: A Decision that Could Threaten the Integrity and Efficiency of WTO Dispute Settlement', *International Economic Law and Policy Blog*, 24 November, <http://worldtradelaw.typepad.com/ielpblog/2015/11/the-tunadolphin-appellate-body-215-ruling-a-decision-that-could-threaten-the-integrity-and-efficiency-of-wto-dispute-settl.html>.
- Hutter, B. (2016), 'A Risk Regulation Perspective on Regulatory Excellence', in Cary Coglianese (ed.), *Achieving Regulatory Excellence*, Washington, DC: Brookings Institution Press.
- Jasanoff, Sheila (1999), 'The Songlines of Risk', *Environmental Values*, 8: 135–152.
- Mavroidis, P. C. (2013), 'Drifting Too Far from Shore: Why the Test for Compliance with the TBT Agreement Developed by the WTO Appellate Body is Wrong, and What Should the AB Have Done Instead', *World Trade Review*, 12: 509–531.
- Paoli, G. and A. Wiles (2015), 'Key Analytical Capabilities of a Best-in-Class Regulator', Penn Program on Regulation Research Paper, <https://www.law.upenn.edu/live/files/4710-paoliwiles-ppr-researchpaper062015pdf>.
- Parker, R. W. (1999), 'The Use and Abuse of Trade Leverage to Protect the Global Commons: What We Can Learn from the Tuna-Dolphin Conflict', *Georgetown International Environmental Law Review*, 12: 1–123.
- Schauer, F. (1991), *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Oxford: Oxford University Press.
- Schauer, F. and R. Zeckhauser (2007), 'Regulation by Generalization', *Regulation & Governance*, 1: 68–87.
- Smith, A. and B. Yandle (2014), *Bootleggers and Baptists: How Economic Forces and Moral Persuasion Interact to Shape Regulatory Policies*, Washington, DC: Cato Institute.