INSTITUTE FOR INTERNATIONAL LAW AND JUSTICE
NEW YORK UNIVERSITY SCHOOL OF LAW

International Law and Justice Working Papers

IILJ Working Paper 2013/4

AN ECONOMIC PERSPECTIVE ON AS SUCH/FACIAL VERSUS AS APPLIED CHALLENGES IN THE WTO AND U.S. CONSTITUTIONAL SYSTEMS

ALAN SYKES

Institute for International Law and Justice
139 MacDougal Street, 3rd Floor
New York, NY 10012
Website: www.iilj.org

ISSN: 1552-6275

Working Papers are issued at the responsibility of their authors, and do not reflect views of NYU, the IILJ, or associated personnel.

New York University School of Law
New York, NY 10012
U.S.A.

Cite as:
IILJ Working Paper 2013/4
(www.iilj.org)

An Economic Perspective on As Such/Facial versus As Applied Challenges in the WTO and U.S. Constitutional Systems

Alan O. Sykes*

Abstract

Suppose that a government governed by some treaty regime enacts a measure that authorizes an official to act contrary to the treaty. Suppose further, however, that no action in violation of the treaty has yet occurred. Should beneficiaries of the treaty obligation be allowed to invoke its dispute resolution process to challenge the measure in advance of such a violation? Or should they be required to wait until a violation has actually occurred? These and related questions raise important issues in the WTO and the U.S. Constitutional systems. This paper employs an economic perspective to address these issues. Among the principal conclusions are that various WTO and U.S. "rules" to govern such matters are foolish from an economic standpoint. Interestingly, however, neither system actually follows its own rules consistently. Adjudicators in both systems have deviated significantly, and many of the reported decisions comport with factors that the economic analysis identifies as important.

Consider a government that is subject to the rules of some treaty regime, and suppose that the government enacts a measure[1] that authorizes some official(s) to act contrary to the treaty. Assume further, however, that no action in violation of the treaty regime has yet occurred. Should other members of the treaty regime (or other actors with standing to enforce the rules) be allowed to invoke its dispute resolution process to challenge the measure in advance of such a violation? Or should they be required to wait until a violation has actually occurred? These and related questions raise important issues

* Robert A. Kindler Professor of Law, NYU Law School. I thank Loic Coutelier for research assistance, and Helen Hershkoff, Michael McConnell, Adam Samaha, participants at the Stanford University Conference on the Future of the WTO in April, 2012, and the participants at the NYU faculty workshop for helpful comments.

[1] The term "measure" should be understood broadly to encompass statutes, regulations, and administrative policies.

in various legal systems, including the two that are the focus of this essay: the WTO and the U.S. Constitutional system.

I begin with terminology. In the WTO system, a challenge prior to the application of a measure in violation of the treaty is termed a challenge to the measure "as such." A challenge only after the measure is applied in a manner that purportedly breaches a treaty obligation is termed a challenge to the measure "as applied."

This terminology is incomplete with respect to the U.S. Constitutional system. In constitutional adjudication, an "as applied" challenge refers to a claim that a measure is unconstitutional if applied under certain factual circumstances. A "facial" challenge involves a claim that the measure is unconstitutional across the board, regardless of the factual circumstances to which it might be applied. It is also useful to distinguish challenges prior to the enforcement of a measure from challenges that occur after enforcement of a measure. Pre-enforcement challenges (the analogue to the as such challenge in the WTO) are often facial, and post-enforcement challenges (the analogue to the as applied challenge in the WTO) are often as applied, but this mapping is not perfect. A pre-enforcement challenge might be "as applied," for example, if the challenge seeks to invalidate a measure prospectively only in so far as it might in the future be applied to particular circumstances. Likewise, a post-enforcement challenge might be termed "facial" if it seeks to invalidate the measure in all possible future applications, and not just under the factual scenario in which it has already been applied.

The issues here relate closely to a number of other legal doctrines that arise in U.S. constitutional (but not WTO) cases. For example, regardless of the type of challenge to a measure, a *remedial* question may arise as to whether the measure can be modified or narrowly construed by an adjudicator to limit its application to circumstances that do not violate the rules. This issue is sometimes couched as a question whether the illegal applications of a measure are *severable* from its permissible applications. In addition, constitutional courts occasionally entertain complaints alleging that a measure is *overbroad*. In effect, the complainant argues that even if the measure as applied to the complainant may be legal, the measure should be deemed

unenforceable across the board because of its possibly illegal applications in other circumstances. The discussion in this paper has bearing on all of these questions.

The paper employs an economic perspective, asking which rules are better or worse from a social welfare standpoint. This paper is among the first to my knowledge to address that issue within a general economic framework.[2] To preview some of the principal conclusions, both the WTO and U.S. Constitutional systems have evolved nominal "rules" to limit as such challenges and facial challenges. In the WTO, the traditional distinction was between "mandatory" and "discretionary" legislation. In the United States, an oft-cited principle is that facial challenges to legislation are allowed only if no set of facts exists under which the legislation would be constitutionally valid (the rule of *United States v. Salerno*[3]). Both of these rules seem foolish from an economic standpoint. Interestingly, however, neither system actually follows its own rules consistently – adjudicators in both systems have deviated significantly. Many of the reported decisions in both systems make considerably more economic sense than the nominal rules.

I. Legal Background: Pertinent Aspects of WTO and U.S. Constitutional Law

The WTO legal system (along with its GATT predecessor) and U.S. Constitutional law each have a substantial body of decisional law regarding the circumstances under which adjudicators will entertain as such or facial challenges. The U.S. Constitutional system has additional layers of jurisprudence concerning pre-enforcement and post-enforcement challenges, remedies, and overbreadth challenges. As shall be seen, the general principles are often subject to important and imprecise exceptions, with the result that the law in both systems is actually rather muddled.

---

[2] The one paper to my knowledge that loosely invokes economic considerations is David H. Gans, Strategic Facial Challenges, Boston University Law Review 85: 1333-88 (2005). Gans argues, among other things, that a strong presumption against facial ('as such") constitutional challenges, suggested by a number of Supreme Court decisions, is unwise because it ignores the potentially higher litigation costs of a series of "as applied" challenges.

[3] 481 U.S. 739 (1987).

A. WTO Law

Challenges to legislation as such arose regularly under GATT (prior to the creation of the WTO). In response, a series of GATT panel decisions evolved what came to be known as the "mandatory/discretionary" distinction. "Mandatory" measures have the property that, in the event of certain factual contingencies, officials are obliged act in a manner that violates GATT. "Discretionary" measures, by contrast, afford the officials who administer them the ability to avoid violating GATT in any contingency that may arise, even if those officials may also possess discretion to act in a manner that contravenes GATT. The general rule under GATT was that mandatory legislation could be challenged as such, while discretionary legislation could only be challenged as applied.

United States – Superfund[4] is illustrative of the distinction. The case concerned taxes to fund an environmental cleanup fund known as Superfund. One part of the challenged legislation imposed an excise tax on imported crude oil higher than the tax imposed on domestic crude oil. Such tax discrimination violates the national treatment obligation of GATT Article III(2), and Mexico brought a case challenging the tax before it went into effect. The panel ruled that the challenge was appropriate:

> [T]he national treatment obligation of Article III…[exists] to protect expectations of the contracting parties as to the competitive relationship between their products and those of the other contracting parties. [Its purpose is] not only to protect current trade but also to create the predictability needed to plan future trade. That objective could not be attained if contracting parties could not challenge existing legislation mandating actions at variance with the General Agreement until the administrative acts implementing it had actually been applied to their trade…[T]he very existence of mandatory legislation providing for an internal tax, without it being applied to a particular imported product, should be regarded as falling within the scope of Article III:2, first sentence. The Panel noted that the tax on certain imported substances had been enacted, that the legislation was mandatory and that the tax authorities had to apply it

[4] United States – Taxes on Petroleum and Related Substances, adopted June 17,1987, 34th Supp. BISD 136 (1988).

> after the end of next year and hence within a time frame within which the trade and investment decisions that could be influenced by the tax are taken.[5]

Another part of the Superfund legislation authorized the imposition of a penalty tax on certain imported substances if importers did not supply certain information to taxing authorities. Such a tax would also run afoul of the national treatment obligation. Because the regulations implementing the system had not been issued, and because the legislation afforded the Secretary of the Treasury sufficient discretion to draft rules to avoid the penalty tax on imports, the panel rejected the challenge to this portion of the legislation:

> From the perspective of the overall objectives of the [GATT] it is regrettable that the Superfund Act explicitly directs the United States tax authorities to impose a tax inconsistent with the national treatment principle but, since the Superfund Act also gives them the possibility to avoid the need to impose that tax by issuing regulations, the existence of the penalty rate provisions as such does not constitute a violation of the United States obligations under the General Agreement.

The mandatory/discretionary distinction was applied strictly and consistently throughout the GATT years. As the last adopted GATT (pre-WTO) panel report on the matter remarked: "[L]egislation which mandated action inconsistent with the General Agreement could be challenged as such, whereas legislation which merely gave the discretion to the executive authority of a contracting party to act inconsistently with the General Agreement could not be challenged as such; only the actual application of such legislation inconsistent with the General Agreement could be subject to challenge."[6]

Early decisions following the creation of the WTO also embraced the mandatory/discretionary distinction,[7] but it soon came under attack to the degree that it implies a total bar to as such challenges of

---

[5] Id. paras. 5.2.1-5.2.2.

[6] United States – Measures Affecting the Importation, Internal Sale and Use of Tobacco, adopted October 4, 1994, 41st Supp. BISD 131, para. 118.

[7] See Argentina – Measures Affecting Imports of Footwear, Textiles, Apparel and Other Items, WT/DS56/R, adopted as modified April 22, 1998, para. 6.45.

discretionary legislation.[8] In *United States – Section 301,*[9] the European Communities challenged provisions in the U.S. Trade Act of 1974 that had been used extensively by the United States during the GATT years to impose retaliation for measures that the United States deemed a violation of its rights under trade agreements. These unilateral determinations by the United States were a major irritant in the GATT system and were an important impetus for the WTO Dispute Settlement Understanding(DSU).[10] Article 23 of the DSU accordingly provides: "Members…shall not make a determination to the effect that a violation has occurred…except through recourse to dispute settlement in accordance with the rules and procedures of this Understanding."

Notwithstanding the DSU, the United States did not (and has not to this day) repealed the pertinent provisions of the 1974 Trade Act. It still allows the United States Trade Representative (USTR) to investigate alleged violations of trade agreements by U.S. trading partners, and sets forth a timeline for USTR to make certain determinations in those investigations, including a determination "whether the rights to which the United States is entitled under any trade agreement are being denied.[11]" The required timeline raises the distinct possibility that USTR might be required to determine "whether" U.S. rights are violated before the WTO dispute settlement process has run its course.

---

[8] These developments spurred a number of doctrinal commentaries and suggestions for reform. See Sharif Bhuiyan, Mandatory and Discretionary Legislation: The Continued Relevance of the Distinction under the WTO, Journal of International Economic Law 5: 571-604 (2002); Kwan Miat Sim, Rethinking the Mandatory/Discretionary Legislation Distinction in WTO Jurisprudence, World Trade Review 2:33-64 (2003); Yoshiko Naiki, The Mandatory/Discretionary Doctrine in WTO Law: The *US – Section 301* Case and its Aftermath, , Journal of International Economic Law 7: 23-72 (2004); In Search of Relevant Discretion: The Role of the Mandatory/Discretionary Distinction in WTO Law, Journal of International Economic Law 13: 379-421 (2010).

[9] United States – Sections 301-310 of the Trade Act of 1974, WT/DS152/R, adopted January 27, 2000.

[10] See Warren F. Schwartz & Alan O. Sykes, The Economic Structure of Renegotiation and Dispute Settlement in the World Trade Organization, 31 Journal of Legal Studies S179 (2002).

[11] 19 U.S.C. §2414(a)(1)(A)(i).

The European Communities brought a case arguing that these provisions of the 1974 Act were mandatory legislation that could be challenged as such. The United States responded that the legislation was discretionary, in that USTR had discretion to determine "whether" U.S. rights had been violated under the statutory timeline, and would exercise that discretion so as to avoid answering the question in the affirmative prior to the conclusion of the WTO dispute settlement process. The panel accepted this interpretation by the United States of its own statute, but further held that even if the legislation was discretionary, an as such challenge was still possible:

> In our view the appropriate method in cases such as this is to examine with care the nature of the WTO obligation at issue and to evaluate the Measure in question in the light of such examination. The question is then whether, on the correct interpretation of the specific WTO obligation at issue, only mandatory or also discretionary national laws are prohibited. We do not accept the legal logic that there has to be one fast and hard rule covering all domestic legislation. After all, is it so implausible that the framers of the WTO Agreement, in their wisdom, would have crafted some obligations which would render illegal even discretionary legislation and crafted other obligations prohibiting only mandatory legislation? Whether or not Section 304 violates Article 23 depends, thus, first and foremost on the precise obligations contained in Article 23.[12]

The panel thus departed from the mandatory/discretionary distinction to hold that an "as such" challenge to discretionary legislation is possible if a proper reading of the substantive obligation at issue implies a prohibition on discretionary measures that might conflict with the obligation. It then went on to rule that Article 23 contains such a prohibition: "[T]he very discretion granted under Section 304, which under the U.S. argument absolves the legislation, is what, in our view, creates the presumptive violation…Members locked in a dispute with the US will be subject to a mandatory determination by USTR under a statute which explicitly puts them in that very danger which Article 23 was intended to remove.[13]" The panel also held, however, that the extensive assurances from the United States that it would not exercise its discretion in a way that violated Article 23 were

---

[12] United States – Section 301, para. 7.53.
[13] Id. para. 7.61.

sufficient to eliminate the "presumptive violation." The panel decision was not appealed.

To date, the WTO Appellate Body has not offered a definitive rule regarding the permissibility of as such challenges. It has, however, addressed the nature of "discretion" in relation to the definition of discretionary legislation. In United States – Antidumping Act of 1916,[14] Japan challenged a little-used 1916 statute on the grounds that it afforded a remedy for dumping beyond that allowed in the WTO Antidumping Agreement. Litigation under the 1916 Act was pending in Federal court against a Japanese company at the time but had not been decided. The United States argued that the language of the Act was sufficiently broad that a court had the discretion to construe it in a way that was consistent with WTO obligations, and thus resisted the as such challenge. The Appellate Body rejected the argument, stating that "in the case law developed under the GATT 1947, the distinction between mandatory and discretionary legislation turns on whether there is relevant discretion vested in the *executive branch* of government.[15]" Because the U.S. position rested on the exercise of judicial discretion, the legislation was not "discretionary" and Japan's challenge was allowed to go forward.

Later, in United States – Section 211,[16] the Appellate Body offered a rationale for the distinction between judicial and legislative discretion: "[W]here discretionary authority is vested in the executive branch of a WTO member it cannot be assumed that the WTO member will fail to implement its obligations under the WTO Agreement in good faith.[17]"

Finally, in *United States – Corrosion Resistant Steel*,[18] the Appellate Body confronted a challenge to the "zeroing" method used by the United

---

[14] WT/DS162/AB/R, adopted September 26, 2000.
[15] Id. Para. 100.
[16] United States – Section 211 Omnibus Appropriations Act of 1998, WT/DS176/AB/R, adopted January 2, 2002.
[17] Id. para. 259.
[18] United States – Sunset Review of Antidumping Measures on Corrosion-Resistant Carbon Steel Flat Products from Japan, WT/DS244/AB/R, adopted January 9, 2004.

States to compute dumping margins.[19] The practice of zeroing was embedded in computer code used by the Department of Commerce in antidumping cases and was reflected in a "policy bulletin" of the Department which provided policy guidance but was not a mandatory regulation. Japan sought to challenge these policies as such. The United States contended that the policies were not challengeable "measures" at all, and argued further that the Commerce Department had legal discretion to depart from its past policies. The Appellate Body rejected both arguments.

In so doing, it offered a policy rationale for as such challenges:

> [T]he disciplines of the GATT and the WTO, as well as the dispute settlement system, are intended to protect not only existing trade but also the security and predictability needed to conduct future trade. This objective would be frustrated if instruments setting out rules or norms inconsistent with a Member's obligations could not be brought before a panel once they have been adopted and irrespective of any particular instance of application of such rules or norms. It would also lead to a multiplicity of litigation if instruments embodying rules or norms could not be challenged as such, but only in the instances of their application.[20]

It went on to hold that the scope of the term "measures" is broad, encompassing "acts setting forth rules or norms that are intended to have general and prospective application," and affirmed the finding of the panel that the U.S. policies amounted to "measures.[21]"

---

[19] Zeroing is a method for the computation of antidumping "margins." To illustrate, suppose that a respondent company makes two sales in the United States market. The "fair value" of the good being sold is $100. One sale is made at a price of $110, and the other at the "dumped" price (a price below fair value) of $90. The average sales price is $100, equal to fair value, and by that reasoning one might say that no dumping has occurred. Using the zeroing method of computation, however the Commerce Department would say that the "margin of dumping" is zero for the sale at $110, and $10 for the sale at $90. Thus, the average margin of dumping would be computed as $5, and a $5 antidumping duty would be imposed on future sales. This method of inflating dumping margins has long been criticized by commentators. See general the papers collected in Richard Boltuck and Robert E. Litan eds., Down in the Dumps (Brookings Institution 1991). In a series of recent WTO decisions, including Corrosion-Resistant Steel, the WTO Appellate Body has ruled that zeroing is inconsistent with the WTO Antidumping Code.

[20] Id. Para. 82.

[21] See also United States – Sunset Reviews of Anti-Dumping Measures on Oil

As for the argument that the "measures" were not mandatory, the Appellate Body remarked:

> [W]e see no reason for concluding that, in principle, non-mandatory measures cannot be challenged "as such"…We observe, too, that allowing measures to be the subject of dispute settlement proceedings, whether or not they are of a mandatory character, is consistent with the comprehensive nature of the right of Members to resort to dispute settlement to "preserve [their] rights and obligations … under the covered agreements, and to clarify the existing provisions of those agreements". As long as a Member respects the principles set forth in Articles 3.7 and 3.10 of the DSU, namely, to exercise their "judgment as to whether action under these procedures would be fruitful" and to engage in dispute settlement in good faith, then that Member is entitled to request a panel to examine measures that the Member considers nullify or impair its benefits. We do not think that panels are obliged, as a preliminary jurisdictional matter, to examine whether the challenged measure is mandatory. This issue is relevant, if at all, only as part of the panel's assessment of whether the measure is, as such, inconsistent with particular obligations.[22]

A similar conclusion was reached in United States – Zeroing (EC).[23] The zeroing method had been applied consistently in a sequence of antidumping cases, even though it was not embodied in a written document.[24] The Appellate Body held that such a consistent practice, even though not expressed in writing, was nonetheless a "measure" that was subject to challenge. Further, it ruled that it could be considered a "rule or norm of general application" based on its repeated use, the fact that no recent antidumping case had been identified in which it was not employed, and the fact that a Commerce Department manual described it as "standard" practice.[25] As a rule or norm of general application, it was properly subject to an as such challenge.[26]

Country Tubular Goods from Argentina, WT/DS268/AB/R, adopted December 17, 2004 (paras. 172-73, 186-87)(affirming that the zeroing procedures embodied in the "Sunset Policy Bulletin" were a "measure" that could be challenged).

[22] United States -- Corrosion-Resistant Steel, para. 89.

[23] United States -- Laws, Regulations and Methodology for Calculating Dumping Margins (Zeroing), WT/DS294/AB/R, adopted May 9, 2006.

[24] These cases were not governed by the written Sunset Policy Bulletin.

[25] Id. paras. 201-205.

[26] Contrast United States – Continued Existence and Application of Zeroing Methodology, WT/DS350/AB/R, adopted February 19, 2009, in which the continued

In sum, the Appellate Body now rejects a strict requirement that measures be mandatory before an as such challenge can go forward. The policies at issue in the zeroing dispute were indeed discretionary policies of executive branch officials, albeit policies that had been applied consistently in the past. The Appellate Body has not, however, offered much guidance as to which discretionary policies may be challenged as such, other than to inquire whether "the measure is, as such, inconsistent with particular obligations" and, in the antidumping context, to suggest that as such challenges are appropriate for "rules and norms of general application."

The distinction offered in the introduction between pre-enforcement and post-enforcement challenges has not yet found its way into WTO decisions, but it is clear that the distinction has practical significance. *United States – Section 301* was a pre-enforcement as such challenge. The United States had not imposed trade sanctions outside the WTO dispute resolution process after the inception of the WTO, and the EC's case was predicated on the claim that the United States would be compelled by its domestic law to do so in the future. By contrast, some of the zeroing cases were post-enforcement as such challenges, in as much as the United States had already applied zeroing in past antidumping investigations and the question was whether it would continue to do pursuant to a general policy that could be challenged with respect to both past and future applications.

With regard to remedial issues, WTO law is relatively simple. Under the WTO Dispute Settlement Understanding, when a panel or the Appellate Body determines that a WTO member has violated WTO law, it shall "recommend that the Member bring the [challenged] measure into conformity" with pertinent WTO obligations. It may also offer some options for doing so,[27] but it has no power to reform or otherwise adopt a limiting construction of the offending measure.

B. U.S Constitutional Law

---

application of zeroing in 18 particular cases was not characterized a general rule or norm given recent changes in U.S. policy, and hence the case was treated as an "as applied" challenge. Id. paras. 180-81.

[27] WTO Dispute Settlement Understanding, Article 19.

U.S. jurisprudence on the issues of interest here is extensive and enormously intricate. A thorough survey of the case law is far beyond the scope of this essay, and I shall focus exclusively on the decisions of the U.S. Supreme Court. Even with this limited focus, it is impossible to be comprehensive, and it is also difficult to pinpoint principles that the Court follows consistently. I will thus limit the discussion to some prominent strands of doctrine and cases that will be of particular interest in the economic discussion to follow.

As a preliminary, note that the remedial powers of U.S. courts in constitutional cases are broader than those of WTO adjudicators. U.S. courts can sometimes reform an unconstitutional statute by giving it a "limiting construction" that bars its application in circumstances where its application would be unconstitutional while allowing it to stand in other scenarios.[28] Such decisions are often favored in situations where the unconstitutional applications of the statute are "severable" from its constitutional applications. Before taking this route, however, Courts will typically ask whether a limiting construction of the statute (severability) is reasonably apparent and administrable, and whether it would clash with the apparent intention of the legislature in enacting the statute. If the answer to the first question is no, or to the second is yes, then the statute may be invalidated in its entirety.[29]

Putting aside the severability issue, the Court has issued a wide range of decisions regarding the permissibility of facial challenges in different areas. Commentators often divide the cases into two broad categories.[30] The first contains First Amendment cases, in which the Court with some regularity allows facial challenges to statutes (Federal or state) that impose unconstitutional restrictions on speech. A prominent illustration is *Board of Airport Commissioners v. Jews for Jesus,*

---

[28] A leading case rejecting an as such challenge on the grounds that the statute was valid under the facts at hand, and any invalid application to a different set of facts could be addressed through a narrower future ruling, is Yazoo & Mississippi Valley RR v. Jackson Vinegar CO., 226 U.S. 217 (1912).

[29] See Richard H. Fallon, Jr., John F. Manning, Daniel J. Meltzer & David L. Shapiro, Hart Wechsler's The Federal Courts and the Federal System 6th ed. 162-65 (Thomson Reuters 2009) (hereafter Hart & Wechsler).

[30] See id. at 165-83.

Inc.,[31] which involved a challenge to a Los Angeles ordinance prohibiting "First Amendment activities" at Los Angeles International Airport. The plaintiff sought to distribute religious literature and was asked to stop under threat of litigation. The Court (unsurprisingly) ruled that the ordinance was considerably overbroad, regulating protected speech, and that no acceptable limiting construction was apparent. Accordingly, the ordinance was invalidated *in toto*.

Jews for Jesus is a nice example of the First Amendment overbreadth cases, in which statutes regulating speech are invalidated in their entirety, even if their application to the plaintiff might be constitutional (an issue not reached in that case).[32] Another example of this phenomenon is the recent decision in *United States v. Stevens*,[33] which invalidated a Federal statute criminalizing depictions of animal cruelty. The Court articulated the standard for facial invalidation in a First Amendment overbreadth case -- the statute can be invalidated if "a substantial number of its applications are unconstitutional, judged in relation to the statute's plainly legitimate sweep.[34]" The Court then held that many depictions of animal cruelty constitute protected speech and struck down the statute without asking whether it was constitutional as applied to the plaintiff.

The substantiality limitation on the overbreadth doctrine suggested by *Stevens* is familiar from prior cases. As an example in which the limitation was not met, *New York v. Ferber*[35] held that a statute prohibiting the distribution of materials depicting sexual activity by children under 16 was not facially invalid even if it might reach some relatively insubstantial amount of protected speech.

The overbreadth cases further underscore the distinction between pre-enforcement and post-enforcement challenges. *Jews for Jesus* was a post-enforcement facial challenge – the ordinance had been

---

[31] 482 U.S. 569 (1987).
[32] See the discussion of overbreadth cases in *Sabri v. United States*, 541 U.S. 600 (2004) at 610.
[33] 559 U.S. 460 (2010).
[34] Id. at 446-47, citing Washington State Grange v. Washington State Republican Party, 552 U.S. 442 (2008).
[35] 458 U.S. 747 (1982).

applied to the plaintiff, and the plaintiff sought to enjoin its enforcement in all circumstances. On other occasions, facial challenges are brought by plaintiffs to whom the measure has not yet been applied, but who claim that the measure should be invalidated in all of its prospective applications.[36]

Turning to cases outside the First Amendment area, conventional wisdom holds that the Court prefers "as applied" challenges.[37] A leading case for this proposition is *United States v. Salerno*,[38] involving what was framed as a facial challenge to pre-trial detention provisions of the Bail Reform Act. In rejecting the challenge, the Court remarked: "A facial challenge to a legislative Act is the most difficult challenge to mount successfully, since the challenger must establish that no set of circumstances exists under which the Act would be valid.[39]" The *Salerno* test is logically equivalent to a requirement that any conceivable "as applied" challenge will succeed.

A similar case is *United States v. Raines*,[40] involving a facial challenge to a statute that allowed the federal government to sue state officials over voting rights violations. The plaintiffs (state officials) challenged the statute on the grounds that it impermissibly attempted to regulate private conduct as well as official conduct, but the Supreme Court rebuffed the challenge, stating that the Court follows two rules: "one, never to anticipate a question of constitutional law in advance of the necessity of deciding it; the other never to formulate a rule of law

---

[36] See, e.g., Brockett v. Spokane Arcades, Inc., 472 U.S. 491 (1985). The plaintiffs brought a challenge to an anti-pornography statute only a few days after its passage, claiming that it was overbroad.

[37] For surveys noting both the conventional wisdom and numerous examples of cases seemingly inconsistent with it, see Michael C. Dorf, Facial Challenges to State and Federal Statutes, 46 Stan. L. Rev. 235 (1994); Richard H. Fallon, Jr., Facts and Fiction About Facial Challenges, California Law Review 99: 915-74 (2011); David H. Gans, Strategic Facial Challenges, Boston University Law Review 85: 1333-88 (2005); Gillian Metzger, Facial and As Applied Challenges Under the Roberts Court, Columbia Public Law and Legal Theory Working Paper 09161 (2009): Roger Pilon, Foreword, Facial v. As Applied Challenges: Does It Matter?, Cato Supreme Court Review 2008-2009: vii-xvii.

[38] 489 U.S. 739 (1987).

[39] Id. at 745.

[40] 362 U.S. 17 (1960).

broader than is required by the precise facts to which it is to be applied.[41]" Likewise, "the delicate power of pronouncing an Act of Congress unconstitutional is not to be exercised with reference to hypothetical cases," and in the event a future case raised constitutional issues based on the application of the Act to particular circumstances, a "limiting construction" could be given to the statute.[42] *Raines* thus suggests a strong preference outside the First Amendment area to avoid striking down laws on the basis of scenarios that have not yet occurred and, in the event of an unconstitutional application of a law, to give the law a narrow construction that forecloses unconstitutional applications while allowing constitutional ones. Once again, such rulings effectively "sever" the unconstitutional features of a law while allowing the remainder of the law to stand.

More recently, *Sabri v. United States*[43] involved a federal criminal case against an individual accused of bribing an official of an agency that receives federal funds. The defendant argued that Congress had exceeded its authority in enacting the statute, an argument that the Court rejected. The opinion went on to comment more generally on facial challenges, stating that "[f]acial adjudication carries too much promise of premature interpretation of statutes on the basis of factually barebones records.[44]"

The Court has also developed a small amount of jurisprudence relating to the choice between pre- and post-enforcement facial challenges, and the possibility of pre-enforcement as applied challenges. *Gonzales v. Carhart*[45] involved a challenge to a Federal statute banning so-called "partial birth" abortions. The statute contained an exception for procedures necessary to protect the life of the mother, but not for procedures otherwise needed to protect the "health" of the mother. The Court rejected a facial challenge to the statute predicated on the absence of a health exception, noting uncertainty as to the likelihood of circumstances in which a threat to the health but not to the life of the mother would arise. The majority further suggested that should

---

[41] Id. at 21.
[42] Id. at 22.
[43] 541 U.S. 600 (2004).
[44] Id. at 609.
[45] 550 U.S. 124 (2007)

circumstances arise where the ban created a threat merely to health, a pre-enforcement challenge to the statute as applied could be brought, which presumably could result in a limiting construction of the statute. A vigorous dissent argued that a pre-enforcement as applied challenge was not a realistic option in a case involving an imminent threat to the health of the mother.

From these cases alone, one might infer that the Supreme Court is generally hostile to facial challenges and rarely entertains them. In fact, however, as various commentators have observed, the Court often strikes down federal and state laws in their entirety, not merely "as applied" to the facts of the case before it. The Court has struck down abortion regulations in their entirety for failure to provide exceptions for medical necessity, even though the plaintiff did not face medical exigency.[46] Numerous additional examples might be offered, some of which will be presented in later sections. One leading commentator contends, on the basis of a survey of decisions in recent Supreme Court terms, that facial challenges are both more common than "as applied" challenges and more likely to succeed.[47] Additional examples will be offered in subsequent sections.

In sum, the Supreme Court's oft-stated skepticism regarding facial challenges is misleading in relation to its actual practice. Despite the suggestion in *Salerno* that facial challenges will be rebuffed unless a challenged measure would be unconstitutional in all circumstances, in practice there are decisions that invalidate laws in their entirety because *some* circumstances exist in which the law is unconstitutional. These cases are not limited to the First Amendment overbreadth arena. But at the same time, numerous examples exist in which facial challenges are rejected, either because some of the potential applications of the statute are constitutional, or because limiting constructions are available that address the instances of unconstitutional application.

II. The Economics of As Such/Facial Challenges

---

[46] See Ayotte v. Planned Parenthood of Northern New England, 546 U.S. 320, 327-31 (2006)(noting prior cases invalidating regulation in its entirety, but preferring to give a limiting construction in the case at bar).

[47] See Fallon (2011) supra.

This section develops several economic considerations that bear on the social welfare implications of allowing as such or facial challenges (for expositional convenience I will simply use the term "as such" in this section). It will suggest that the economic costs and benefits turn on issues that are poorly captured by the existing legal "rules" in the WTO and U.S. Constitutional systems, and only partially captured by case law developments in each system.

The relevance of economic considerations to the choice of a legal standard for as such challenges will of course be somewhat controversial. My view is that economic considerations are primarily useful for the purpose of fashioning a "default" rule. If a treaty regime includes a clear and specific rule allowing or disallowing such challenges, it should be followed. But if, as in the WTO and U.S. Constitutional systems, the treaty is (arguably) silent on the matter and the issue is left to adjudicators, the costs and benefits of as such challenges are appropriate for adjudicators to consider in deciding whether to allow them.

A. The Simple Analytics of the Choice Between Pre-Enforcement As Such Challenges and Post-Enforcement As Applied Challenges

To introduce the economic issues, let us put aside the complications in the U.S. Constitutional system associated with pre-enforcement with post-enforcement challenges. I will return to these possibilities later. For now, the "as applied" challenge is assumed to occur after a measure has been enforced in a manner that violates the treaty system; an "as such" challenge occurs before enforcement and a violation arise. This dichotomy captures the choice that is made in the WTO system and, as shall be seen, has important implications for the U.S. Constitutional system as well.

On these assumptions, consider a potential dispute between a claimant and a respondent. Assuming that the respondent will abide by the ruling in the case, a successful as such challenge permits a complainant who may suffer future harm to avert the harm altogether. A complainant who is restricted to an as applied challenge, by contrast, must wait until the harm occurs and then seek whatever redress may be

available. The former option may, depending on the circumstances, be quite beneficial to the complainant.

From the respondent's perspective, and again assuming that it will abide by the ruling in the case, a respondent subject to a successful as such challenge must forego the opportunity to take action under the challenged measure in all scenarios, whether or not the exercise of that authority would violate the treaty regime. Thus, to the degree that a measure facilitates conduct that comports with the law, a respondent who loses an as such challenge must incur the costs of fashioning a new measure to facilitate that conduct or else forego that conduct. If measures are subject only to as applied challenges, by contrast, a respondent can take permissible actions pursuant to them without limitation. Should a measure be applied in an illegal manner at some point, the respondent will then become subject to a claim for redress.

The net economic effects of as such challenges will turn on the balance between the benefits to complainants and the costs to respondents. The remainder of this section details a number of considerations that bear on this balance.

1. Necessary Condition for the Efficiency of an As Such Challenge: Complainant's *Ex Post* Remedy is Inadequate

Suppose that in the event of (at least any inefficient[48]) harm to a potential claimant, the claimant will have an as applied challenge and a remedy that is fully and accurately compensatory, in the sense that it restores the welfare of the claimant to its level before the occurrence of the harm. Assume further that no third parties are affected by the potential respondent's conduct. Under these assumptions, as such challenges are simply unnecessary and can easily prove counterproductive. The potential complainant by assumption is insulated from harm, the respondent will internalize costs of inefficient

[48] It is well known that efficient behavior by "injurers" (the analog to the respondent here) can be induced through appropriately crafted strict liability rules (that require compensation for all harm) and negligence rules (that require compensation for inefficient harm). E.g., Steven Shavell, Economic Analysis of Accident Law, Chapter 2 (Harvard Press 1987). The discussion here abstracts from the choice of regime in this respect.

harm to the complainant, and the respondent will be induced to act efficiently.[49] If a claimant can bring an as such challenge nevertheless, it will simply impose unnecessary costs on the respondent, and a claimant may even pursue a challenge strategically to try and extract surplus.

Putting aside third party externalities, these observations suggest that as such challenges are undesirable if claimants can bring successful as applied challenges following inefficient harm and receive full compensation (ignoring litigation cost complications for the moment). Conversely, inadequacy of the *ex post* remedy is a necessary condition for as such challenges to become desirable from an economic standpoint although, as shall be seen, it is by no means sufficient.

An injured party's remedies may prove inadequate for three types of reasons. First, the substantive law at issue may afford no remedy for inefficient harms. For example, imagine a system of tort law under which no injury is compensable unless it is intentionally inflicted. Under these circumstances, the incentive of injurers to take economically worthwhile precautions against accidental harm will be lost, even if the remedy for intentionally inflicted harms is fully compensatory.

Although inefficiencies in the substantive law represent an important class of problems in many fields, they afford little basis for choosing between as such and as applied challenges. If the underlying substantive law fails to condemn inefficient behavior (or prevents efficient behavior), it seems unlikely that either type of challenge can promote efficiency. Accordingly, I will assume that the underlying substantive law is efficient, in the sense that it at least allows a claim for relief whenever the complainant suffers inefficient harm.

A second possible reason for inadequate remedies is legal error in *ex post* adjudication. The complainant with a meritorious as applied

---

[49] If potential claimants are insured against harm, of course, their own incentives to avoid harm may become inefficient as noted by Ronald Coase, The Problem of Social Cost, Journal of Law and Economics 3: 1-44 (1960). To address this problem liability rules may require modification (as through an appropriate affirmative defenses, decoupling of damages paid and received, and the like). Once again the discussion here abstracts from these issues.

challenge may be denied relief by mistake. I will not dwell on this possibility either, however, because it seems unlikely to afford a compelling case for as such challenges. Indeed, for reasons that are developed later, a pre-enforcement as such challenge may be more likely to result in error than a post-enforcement as applied challenge.

The third reason why remedies may prove inadequate is simply that they may fail to afford full and accurate compensation for harm suffered. With particular reference to the WTO and the U.S. Constitutional systems, serious concerns arise in this regard.

Under WTO law, a violator incurs no formal sanction until a complaining member has brought a case, received a favourable adjudication, and the violator has exhausted a "reasonable period of time" to bring its behavior into compliance.[50] The practical result is that a violator can break the rules for a period of years before any formal sanction is triggered, and indeed can avoid any formal sanction completely by curing the violation within the "reasonable period." Commentators sometimes refer to this system as the "three-year free pass." In addition, it may be doubted that the formal remedy for WTO violations after the expiration of the "reasonable period" – trade sanctions imposed by the complainant – can compensate complainants even for the prospective harm suffered due to the ongoing violation.[51] Small countries, for example, cannot use trade sanctions to improve their terms of trade, and often complain that retaliatory measures amount to "shooting themselves in the foot." Larger countries with the ability to improve their terms of trade through sanctions may also be undercompensated for prospective harm because of the principle that trade sanctions must be "equivalent" to the harm caused by the violation, a vague standard administered in somewhat unclear fashion

---

[50] See WTO Dispute Settlement Understanding, Articles 21-22.

[51] To be sure, informal sanctions may exist in addition to formal remedies, and the question whether WTO members can cheat without paying an appropriate price is academically controversial. On the subtleties of whether trade retaliation can compensate trading nations for the welfare losses associated with breach of trade agreements, see Kyle Bagwell, Remedies in the WTO: An Economic Perspective, in M. E. Janow, V. J. Donaldson, and A. Yanovich (eds.), The WTO: Governance, Dispute Settlement and Developing Countries, pp. 733–770; Gene M. Grossman & Alan O. Sykes, Optimal Retaliation in the WTO – A Commentary on the *Upland Cotton* Arbitration, 10 *World Trade Review* 133 (2011).

by WTO arbitrators. If the arbitrators allow only the level of retaliation that restores the complainant's terms of trade, for example, then the complainant is undercompensated because of the decline in trade volume due to the violation.[52]

In the U.S. Constitutional system, remedies are also limited in many cases. Depending on the particular constitutional violation in question, damages for past harm suffered may not be available at all, and the remedy may be limited to an order directing the government to desist from the conduct in question going forward (such as an order declaring that the enforcement of a statute against the plaintiff was unconstitutional). Likewise, some violations may involve conduct that irreparably alters the future course of affairs, such as restrictions on speech or voting that affect the outcome of an election. Measures to restore the *status quo ante* may as a practical matter prove infeasible.

Such limitations on *ex post* remedies in both legal systems raise the distinct possibility that respondents can often behave inefficiently without compensating complainants. As a result, as applied challenges may be inadequate to discourage inefficient behaviour even if the underlying rules prohibit it, and as such challenges may prove useful.[53]

This observation immediately suggests that improvement in *ex post* remedies may be a substitute for allowing as such challenges. In the WTO, retaliation rights might be altered to ensure that complainants are fully compensated for all harm suffered since the inception of a violation (a daunting measurement issue to be sure, which would also likely require compensation to take forms other than trade retaliation, but certainly possible in principle[54]). In the U.S. constitutional system,

---

[52] See Bagwell, supra; Grossman & Sykes, supra.

[53] A loose analogy may be drawn to the literature on the choice between *ex ante* safety regulation and *ex post* liability. A well-known proposition in this regard is that *ex ante* regulation (the loose analogue to an as such challenge) may be preferable when injurers lack the resources to pay *ex post* liability judgments. See Shavell, supra note X, §12.2. Here, a similar problem arises due to the inadequacy of the *ex post* remedy.

[54] Various proposals have been advanced for the use of money damages in the WTO, for example, the most recent being Marco Bronckers & Freya Baetens, Reconsidering Financial Remedies in WTO Dispute Settlement, 16 J. Int'l Econ. L. 281 (2013).

remedies might be expanded to include full damages for any harms unconstitutionally inflicted and so on. In the analysis to follow, however, I will assume that such options are not available for some political or practical reason, and that complainants are limited to an inadequate *ex post* remedy following an as applied challenge.

2. The Basic Cost-Benefit Trade-off

In keeping with the discussion above, assume that the *ex post* remedy for inefficient harm caused to the complainant by the respondent is undercompensatory. The substantive law of the treaty regime governing the conduct of a potential respondent is otherwise assumed to be efficient. Accordingly, behavior that complies with the treaty yields a non-negative net social benefit, and non-compliant behavior yields a non-negative net social cost.

To make it more concrete, suppose that the remedy in the event of a successful complaint (whether as such or as applied) is simply a declaration regarding the legality of the respondent's conduct. Assume further that the respondent will comply with these declaratory rulings and desist from behavior that has been adjudicated to violate the treaty. This assumption arguably captures stylized facts about the WTO and U.S. constitutional systems. In the WTO, for example, members eventually comply with the vast majority of rulings, whether they involve as such or as applied challenges. Yet, because of the "three-year free pass," members arguably take advantage of the system to engage temporarily in behavior that violates the rules.[55] Similarly, in the U.S. Constitutional system, public officials may be overwhelmingly likely to comply with a judicial ruling declaring some action to be unconstitutional, yet the prospect of such a ruling may be insufficient to discourage unconstitutional action in the first place.

Following these assumptions, let the respondent have a domestic measure in place authorizing behavior that may or may not be compliant with the treaty. *Assume for now that the measure will be applied only once*, and that the probability of non-compliant application

[55] See Michael J. Trebilcock & Robert Howse, The Regulation of International Trade (3d ed.) 2005 -146-47.

is *p*. This positive probability of non-compliant (socially inefficient) application arises precisely because the complainant's *ex post* remedy is inadequate and thus fails to deter inefficient behavior by respondent.

Let N denote the net social cost of non-compliant application. The expected social harm if the complainant is restricted to an as applied challenge is thus pN.

On the other side of the ledger, let C denote the net social benefit of compliant application. One can alternatively think of C as the cost to the respondent of enacting a new measure to facilitate compliant behavior in the event that the existing measure is invalidated in an as such challenge (the respondent will incur that cost if it is smaller than the cost of foregoing the benefits of compliant behavior altogether). If the measure is invalidated by an as such challenge, the expected social harm is (1-p)C.

Under these assumptions, the following condition must hold if the as such challenge is to be weakly efficient:

$$pN \geq (1-p)C$$

This simple inequality directs attention to three key parameters. Invalidation of the measure through an as such challenge is more likely to prove efficient as (a) the probability of non-compliant application increases; (b) the harm due to non-compliant application increases; and (c) the benefits of compliant application (or the costs of enacting a new measure to facilitate compliant behavior) decrease.[56]

[56] The reader may notice a loose analogy to Judge Posner's treatment of the standard for granting a preliminary injunction. He suggests that the court should grant a preliminary injunction if the probability that the plaintiff will ultimately prevail times the "irreparable harm" suffered by the plaintiff absent an injunction is greater than the probability that the defendant will ultimately prevail times the "irreparable harm" that the defendant will suffer if the injunction is granted. Richard Posner, Economic Analysis of Law, 5th ed. §21.4 (Little Brown 1998). See also his opinion in Roland Machinery Co. v. Dresser Industries, Inc., 749 F.2d 380 (7th Cir. 1984).

Although omitted from the parameterization above, it is also readily apparent that the adequacy of the complainant's *ex post* remedy is relevant. As the *ex post* remedy improves, the respondent will be less likely to engage in non-compliant (inefficient) application of the measure. Thus, *p* declines and the as such challenge tends to become less useful.

3. Legal Implications

These observations cast immediate doubt on some of the prominent purported "rules" governing as such challenges in the WTO and U.S. Constitutional systems. For example, the *Salerno* rule in U.S. Constitutional law, which holds that facial challenges to U.S. statutes are not allowed unless there is "no set of circumstances" in which the statute would be constitutional, seems clearly misguided from an economic standpoint. It amounts to a requirement that p=1, yet as such challenges can prove efficient under a far broader set of circumstances.[57]

The mandatory/discretionary distinction in WTO/GATT law fares somewhat better, but is still questionable. Perhaps one might argue that the probability of non-compliant application is systematically higher with mandatory legislation, in that adjudicators can imagine factual scenarios in which a violation would certainly occur under the measure. With discretionary legislation, by contrast, there is no set of circumstances in which non-compliant application is compelled by the measure.

This defense of the mandatory/discretionary distinction, however, is not convincing. The probability of non-compliant application associated with mandatory legislation depends on several other factors, including the probability that circumstances will arise in which non-compliant application is compelled. The probability that a mandatory measure will be repealed before it is applied in a non-

[57] Interestingly, in a memorandum supporting the denial of certiorari in Janklow v. Planned Parenthood, 517 U.S. 1174 (1996), Justice Stevens seemed to acknowledge the dubiousness of the *Salerno* rule, characterizing it as imprudent "dictum" and a "rigid and unwise." But no case to date has specifically repudiated the oft-cited *Salerno* principle outside the First Amendment area.

compliant way must also be considered. Likewise, discretionary legislation may have a high probability of non-compliant application even though administrators are legally empowered to avoid it.

Furthermore, the probability of non-compliant application is but one of the three pertinent parameters in this simple framework. None of the "rules" in either system pay explicit attention to the costs of the potentially non-compliant application that will arise if an as such challenge is barred. And none of the rules inquire as to the burden on the respondent of invalidating a measure that facilitates potentially compliant conduct.

The weaknesses of the mandatory/discretionary distinction are readily apparent in a case like Brazil – Export Financing Programme for Aircraft.[58] Brazil was found to be in violation of the Agreement on Subsidies and Countervailing Measures because of its PROEX program, which conferred prohibited export subsidies. Brazil then enacted PROEX III, a new programme, which Canada contended was designed to perpetuate Brazil's prohibited policies even though nominally affording executive discretion to avoid such a violation. The compliance panel rejected Canada's challenge on the basis of the mandatory/discretionary distinction:

> In our view, a conclusion that PROEX III *could* be applied in a manner which confers a benefit, or even that it was intended to be and *most likely would* be applied in such a manner, would not be a sufficient basis to conclude that PROEX III as such is mandatory legislation susceptible of inconsistency with Article 3.1(a) of the SCM Agreement.[59]

In effect, the panel held that even if the Brazilian program was a ruse to maintain illegal policies under the guise of executive discretion to avoid them, an as such challenge was unavailable.

The deficiencies of the purported simple "rules" of each system may explain why adjudicators have proven reluctant to adhere to them in a number of cases. Some of the decided cases in fact take a more

---

[58] Brazil – Export Financing Programme for Aircraft, Second Recourse by Canada to Article 21.5 of the DSU, WT/DS46/RW2, adopted on August 23, 2001.
[59] Id. para. 5.43.

nuanced approach that may better comport with economic considerations, even if adjudicators do not appeal to them expressly. For example, in the *US -- Section 301* case, the WTO panel rejected the notion that discretionary legislation cannot be challenged as such, but ultimately sided with the United States on the merits of the challenge following strong assurances from U.S. officials that they would exercise their discretion to ensure U.S. compliance with the DSU. One can interpret that decision as rejecting a strict bar to as such challenges of discretionary legislation, coupled with a willingness to reject such challenges nonetheless if the probability of non-compliant application appears low.

Similarly, in *US – Corrosion Resistant Steel,* the Appellate Body approved an as such challenge to the zeroing practice of the Department of Commerce despite the fact that the measure in question was simply an administrative practice embedded in departmental computer code and in a non-binding policy bulletin. Despite its discretionary nature in this sense, the practice had been applied consistently in the past in a manner that was found to violate the Antidumping Code, and there was no reason to think that it would not be so applied in the future. Moreover, the burden to the United States of changing its measure was trivial – all that was required was a change in administrative practice and a little bit of computer reprogramming. In this instance, therefore, the probability of non-compliant application was high, and the burden to the respondent of replacing the measure with a compliant one was negligible.

Other decisions are a bit harder to square with the economic framework. In *US – Antidumping Act of 1916*,[60] the Appellate Body rejected the assertion that the legislation was discretionary, and thus inappropriate for an as such challenge, on the grounds that discretion as to its interpretation rested with the judicial rather than the executive branch of government. That distinction seems indefensible. If the language of a statute can be interpreted to accord with WTO obligations, the fact that the interpretive exercise will occur in a court rather than an administrative agency seems irrelevant. The irrelevance of the distinction seems particularly clear where, as in the United States,

[60] WTO/DS162/AB/R, adopted September 26, 2000, para. 100.

principles of statutory construction favor interpretation of domestic laws in a manner that comports with international obligations.[61]

U.S. constitutional jurisprudence has also evolved away from the *Salerno* rule in a number of recent cases, in ways that suggest the implicit relevance of the economic considerations developed above. *Santa Fe Independent School District v. Doe*[62] offers an interesting example. The case involved an Establishment Clause challenge to a school district policy that permitted students to elect a representative to give an "invocation" at high school football games, which could include a religious prayer. The policy was enacted following a lawsuit to challenge a practice whereby an elected "student chaplain" would routinely deliver a prayer at home football games. The plaintiff claimed, in essence, that the new policy was a ruse to circumvent the illegality of the prior practice. The six-vote majority held the new policy to be a violation of the Establishment Clause even though it had not yet been put into effect, over a dissent by Justices Rehnquist, Scalia and Thomas, who insisted that a facial challenge was inappropriate under the *Salerno* rule. They argued, among other things, that until the policy was in force, it could not be known whether the elected student representative would be a religious person or inclined to deliver a religious prayer. The rejection of this argument by the majority can be interpreted as implicit recognition that given the context in which the new policy was enacted, the probability that it would result in routine prayers at football games was high.

City of Chicago v. Morales[63] was another case in which the majority and dissent battled over a facial (Due Process) challenge, in this case to a Chicago Gang Congregation Ordinance. The ordinance prohibited "criminal street gang members" from "loitering" in a public place. Police officers observing such behavior would issue a "dispersal order," and anyone disobeying the order could be arrested and prosecuted. The majority held the ordinance invalid on grounds of vagueness – the term "loitering" did not afford sufficient clarity as to what conduct was

---

[61] The genesis of this principle in U.S. law is Murray v. The Charming Betsey, 6 U.S. 64 (1804).
[62] 530 U.S. 290 (2000).
[63] 527 U.S. 41 (1998).

prohibited, and afforded police officers too much discretion to interfere in protected conduct.

The dissents by Justices Rehnquist, Scalia and Thomas emphasized that there was no evidence that Chicago police officers had misused their discretion to issue inappropriate dispersal orders, and that the convicted respondents were not engaged in any plausibly protected conduct at the time that they violated dispersal orders. They also questioned the importance of "loitering" as protected conduct. Justice Scalia further objected to the invalidation of the ordinance across the board when it was doubtful that a successful as applied challenge could succeed given the facts of the case. He suggested that the majority effectively extended the overbreadth doctrine beyond the First Amendment context and, invoking the *Salerno* principle, argued that facial challenges should not succeed on grounds of vagueness.

The dissenters may be characterized as opposing the facial challenge in a case where it seems suspect from an economic standpoint. They questioned whether the magnitude of the harm due to any non-compliant conduct was serious, and further doubted that the probability of non-compliant conduct was substantial. I take no position on whether they were right as an empirical matter, but their attention to such considerations seems appropriate.

On a final note, consider the First Amendment overbreadth doctrine. Three characteristics warrant explanation: (i) the fact that the Court seems more open to facial challenges (and perhaps total invalidation) in the First Amendment setting than elsewhere; (ii) the existence of the substantiality requirement noted in cases such as *United States v. Stevens* and *New York v. Ferber*; and (iii) the possibility (evident in *Jews for Jesus* or *Stevens*) that a plaintiff will prevail in an overbreadth challenge, even if the measure at issue may be constitutional as applied to the plaintiff, because of its possibly unconstitutional applications to other potential plaintiffs.

The first characteristic might seem explicable by the Court's observation in *Gooding v. Wilson* that First Amendment expression has

"transcendent value to all society.[64]" Within the analytic framework above, this proposition suggests that the costs of non-compliant application (N) are especially high in the First Amendment setting. As a general proposition, however, this claim seems somewhat dubious – the "transcendent value" of depictions of animal cruelty, for example, or religious proselytizing in airports, is not entirely obvious. Perhaps a further explanation lies in the fact that opportunities for speech are often fleeting and irretrievable. When a statute unconstitutionally discourages valuable speech over a period of time and space, the difficulties of remedying the harm *ex post* are acute. Damages for past harm are generally unavailable. And as noted, the weaker the *ex post* remedy, the more useful a facial challenge may become.

The substantiality requirement fits neatly within the analytic framework above. It amounts to a requirement that *p* be relatively substantial [and hence that (1-p) be relatively less substantial]. Other things being equal, this condition tends to enhance the likelihood that facial challenges are efficient.

Finally, consider the possibility that a plaintiff may prevail in an overbreadth challenge against a measure that is constitutional as applied to the plaintiff, but unconstitutional in other settings. A possible justification relates to the observations made in connection with point (i) above. If the remedy for retrospective harm in the First Amendment setting is negligible, and a plaintiff can only obtain a remedy that facilitates future speech, the likelihood of litigation against unconstitutional infringements of the First Amendment may often be low. It may then make sense to allow a plaintiff, against whom a measure may be applied constitutionally, to assert the rights of parties not before the Court who may suffer harm due to unconstitutional applications of the measure. Plaintiffs in the former group will not sue, however, if they do not receive something valuable by way of remedy. It is thus necessary to invalidate the measure as to them to encourage them to bring suit against measures that infringe the rights of others.

B. Extensions and Complications

[64] 405 U.S. 518, 522 (1972).

The simple framework above assumes away a number of factors that also bear on the economic effects of as such challenges. This section briefly considers several extensions to the analysis.

Litigation Costs and Serial As Applied Litigation. I assumed above that each measure would be applied exactly once, but of course most measures can apply in a number of cases. This extension can be accommodated to a degree by simply interpreting p and (1-p) above as capturing the percentage of cases involving non-compliant application and compliant application, respectively. The interpretation of N and C can be modified accordingly as well.

This simple reinterpretation of parameters, however, masks an important practical consideration. In both the WTO dispute settlement mechanism and the U.S. judicial system, litigants do not bear the full cost of the adjudication process. Not only are the direct costs of adjudication subsidized, but the costs to other potential litigants of congestion in the process are not borne by litigants.

Attention to litigation and congestion costs raises a number of complications, and I will mention just a few. First, if complainants are restricted to as applied challenges, will they be forced to litigate seriatim in a large number of costly proceedings to vindicate their rights?[65] If so, the litigation costs associated with as applied challenges may become quite large (a larger value of N), and as such challenges may have appeal as a way to put the problem to rest in a single proceeding. An analogy exists in the zeroing litigation against the United States under WTO law, where a large number of cases have been brought challenging zeroing in a variety of legal contexts,[66] all of which have resulted in a finding that zeroing violates the WTO Antidumping Code. A single proceeding addressing zeroing in all cases would likely have been much cheaper.

---

[65] This is the primary concern of Gans, supra.

[66] The WTO website lists 17 separate cases involving the U.S. practice of zeroing filed between 1998 and 2012. See http://www.wto.org/english/tratop_e/dispu_e/dispu_subjects_index_e.htm#selected_subject

The additional litigation costs that may be associated with as applied challenges will depend importantly on the precedential status of as applied challenges and the willingness of the respondent to honor precedent without further litigation. Conceivably, a single as applied challenge may establish a precedent that the respondent will respect going forward, and that will suffice to prevent any further non-compliant applications of the measure at issue. Then, a single as applied challenge serves as well as a single as such challenge at eliminating non-compliant behavior. But if the respondent does not conform its behavior to as applied precedents, or if the facts of each case differ sufficiently for the respondent to argue that prior as applied precedents are not on point with respect to future applications of a measure (as in the zeroing cases[67]), then as such challenges may yield considerable reductions in litigation cost.

But there are countervailing considerations. The possibility arises that as such challenges will be pursued excessively to ward off conceivable but unlikely non-compliant behavior. One can understand the reluctance of U.S. courts to issue "advisory opinions" before a clear controversy arises in part as an effort to avoid situations in which litigants take advantage of subsidized adjudication to obtain rulings on hypothetical disputes that will never arise, imposing congestion costs on other litigants in the process.

In short, as such challenges can be either a blessing or a curse when litigation costs are factored into consideration. Nevertheless, however, the factors identified in the last section remain a key focus. If serial as applied litigation is likely, the situation conforms roughly to one in which the probability of non-compliant behavior is high and as such challenges are more likely to be useful. Where the complainant is seeking an "advisory opinion" relating to an unlikely future contingency, p is small, and the as such challenge is less desirable.

Remedy Redux. As noted, as such challenges are unnecessary if respondents must fully compensate complainants following meritorious as applied challenges, and as such challenges tend to become less

[67] In the various zeroing cases, the United States took the position that the rules on zeroing in antidumping cases differed across original investigations, administrative reviews, sunset reviews, and new shipper investigations.

attractive as the *ex post* compensation to respondents approaches full compensation (because the probability of non-compliant application diminishes). Under these circumstances, the decision by a respondent to retain a measure that permits non-compliant behavior may suggest that the costs to the respondent of fashioning a new measure to facilitate compliant behavior are high. Accordingly, even if *ex post* compensation is somewhat inadequate, an argument can sometimes be made for rejecting as such challenges as long as compensation is close to satisfactory.

A parallel set of issues relates to the adequacy of the remedy in the event of a successful as such challenge. The analysis above assumes that the respondent will comply with an adverse ruling in an as such case. This is a plausible assumption to make given the compliance rate in WTO and U.S. Constitutional cases. In the WTO system, however, note that the formal sanction for non-compliance with a ruling (following the reasonable period allowed) is a trade sanction "equivalent to" the harm imposed by the violation under DSU Art. 22(4). If a measure is held to violate the law "as such" before it has been applied, however, one might argue that no harm has occurred and hence there can be no formal sanction (but see the section below on "chilling effects"). This argument has yet to be tested in arbitration over trade sanctions, but if it is correct, a danger arises that successful as such challenges might at some future time be followed by systematic non-compliance with them. Such a development might prove unfortunate for the stability of the WTO system, and perhaps offers a caution about entertaining as such challenges before complainants have suffered any demonstrable injury.

Severability. The discussion above assumes that a successful as such challenge results in total invalidation of the measure at issue. The respondent loses the benefit of engaging in compliant behavior under the measure, and must forego those benefits or incur the costs of formulating a new measure to realize them.

As noted earlier, however, adjudicators may have the option of "reforming" the measure in question. For example, an order might issue directing that the measure not be applied in certain ways that violate treaty obligations, but permitting it to apply in other ways. Similarly, in

the case of a measure that contains ambiguity, adjudicators might resolve the ambiguity by giving the measure an interpretation that comports with treaty obligations. Finally, if a measure contains distinct parts, some of which are compliant and others non-compliant, the non-compliant part of the measure can perhaps be "severed" from the remainder, allowing the compliant behavior authorized by the measure to continue. I will refer to all these possibilities as the "severability" option.

The severability option can, in principle, be invoked following either an as applied or an as such challenge. If severability is not in tension with the wishes of the actors who created the measure (discussed below), it generally dominates an as such adjudication that invalidates a measure in its entirety, as it avoids the costs of formulating an alternative measure to facilitate compliant behavior while addressing the problem of non-compliant behavior.

This proposition is subject to an important caveat: a question may arise whether a measure stripped of its non-compliant elements may be worse from a social standpoint than no measure at all. This issue is often framed as the question whether the reformed measure would be consistent with the intent of its drafters. Precisely that question arose in U.S. Constitutional litigation over the Patient Protection and Affordable Care Act.[68] If the individual insurance purchase mandate had been deemed unconstitutional, for example, would Congress have wanted the prohibition on insurance policies containing pre-existing condition exclusions to stand? Even if not, would it have wanted the rest of the Act to remain in effect? Clearly, the severability option is unappealing if modification of a measure is inferior to its abolition.

A further difficulty with severability can arise when it is difficult to fashion a workable rule that distinguishes compliant from non-compliant application. *Jews for Jesus*, discussed previously, is a nice illustration, where the Court found it infeasible to devise a reasonable limiting construction of an ordinance prohibiting "First Amendment activities" at LAX.

---

[68] See National Federation of Independent Business v. Sebelius, 567 U.S. __ (2012).

Finally, severability is of no value unless adjudicators are empowered to order it. Although U.S, courts exercise such power in various contexts, WTO adjudicators lack any authority to modify the domestic laws of WTO members. They simply recommend that members bring their measures into compliance, leaving to members the question of exactly how to do so in a given case (although they may "suggest" some options pursuant to DSU Article 19(1), as noted).

Conceivably, although panels and the Appellate Body have not done so to date, they might issue "conditional" decisions following as such challenges, holding that certain conduct pursuant to a challenged measure *would* violate WTO obligations. Such decisions, if respected by the respondent, might serve to prevent non-compliant behavior before it occurs. They might also afford a basis for expedited proceedings in the future if a respondent proceeded to apply the measure in a non-compliant way (perhaps the complainant could immediately proceed to an Article 21.5 compliance panel). Nothing in the DSU clearly prohibits such an approach, and even if provisions such as Article 19 were interpreted as limiting adjudicators to unconditional findings that measures do or do not conflict with WTO obligations, the possibility of conditional decisions might be a useful reform to consider.

Pre-Enforcement and Post-Enforcement Adjudication. The analytic discussion in Section II.A embraced the simplifying assumption that as applied challenges occur post-enforcement, while as such challenges occur pre-enforcement. These assumptions are accurate for the WTO and in many U.S. cases, but as noted previously it is possible to imagine pre-enforcement as applied challenges, as well as post-enforcement as such challenges.

As an example of a possible pre-enforcement as applied challenge, recall the decision in *Gonzales v. Carhart,* which declined to invalidate on its face a ban on "partial birth abortions" for lack of an exception for procedures needed to protect the health of the mother. In so doing, the Court suggested that should such a scenario arise in the future, an as applied challenge could be brought (and presumably adjudicated in time for the mother to have any necessary procedure). The dissent

questioned, among other things, whether adjudication could proceed quickly enough.

Assuming that adjudication can occur with sufficient rapidity to avert harm, a pre-enforcement as applied challenge has much virtue. By hypothesis, it avoids the net social harm from non-compliant application of a measure, but it also avoids the social harm from invalidating a measure as to its compliant applications. In this sense it is the best of all worlds. The difficulty, of course, is that as applied adjudication as to merely threatened harm is often infeasible as a practical matter. WTO cases routinely consume two to three years, and Federal court litigation in the United States can move notoriously slowly. One can imagine efforts by litigants and courts to accelerate adjudication in some of these cases, or perhaps changes in procedural rules to facilitate quicker adjudication, but given the overall limitations on judicial resources, any shift of resources toward this class of cases will likely have opportunity costs associated with greater delays elsewhere.

Now consider post-enforcement as such decisions. Examples abound, such as many of the zeroing cases in the WTO system. These cases raise no new issues from an economic standpoint. The complainant seeks a broad ruling invalidating the measure at issue in all future applications, typically (as in the zeroing cases) to ensure that the complainant does not have to pursue a future case concerning future applications of the measure. The costs of serial as applied litigation, discussed earlier, are the driving force here. The disadvantage of an as such ruling is as before – if a measure has important compliant applications, the respondent must either forego them or incur the costs of fashioning a new measure to facilitate them. The balance of these competing considerations will determine whether an as such ruling for the complainant is efficient.

Chilling Effects. The discussion above proceeds on the assumption that compliant behavior yields a net social benefit, while non-compliant behavior yields a net social cost. One might also imagine situations in which the mere *possibility* of non-compliant behavior yields a social cost. In more conventional parlance, perhaps a possibility of non-

compliant behavior creates a chilling effect on permissible and socially beneficial conduct.

Concern for a chilling effect is familiar in U.S. Constitutional jurisprudence, particularly with respect to restrictions on speech challenged under the First Amendment. Unduly vague prohibitions on "obscenity" might discourage constructive artistic expression for fear of prosecution;[69] unduly lax standards for lawsuits claiming libel and slander might discourage critical reporting about important individuals and events.[70]

A similar concern may arise with international trade. If all trade transactions were instantaneous and entailed no sunk investments, potential future treaty violations would have no impact on current transactors. Any harm from such violations would await the time at which the violation occurs and transactions are disrupted. But suppose that exporters must make sunk investments to develop customer- or country-specific trading relationships. Perhaps goods must be customized for particular markets and customers, perhaps dedicated sales forces must be hired and trained, perhaps dedicated production or distribution facilities must be built, and so on.[71] Suppose further that the expected returns to such investments are reduced by a possibility of some breach of WTO rules. The possibility thus arises that efficient investments will be discouraged by measures that create the mere possibility of a future treaty violation.

Of course, investments may be discouraged by all manner of uncertainties about the future – risks of political instability, risks of new regulatory policies, risks of takings for public use, risks of legal or illegal trade barriers, and the like. In general, it is efficient for investors to take

---

[69] See, e.g., Reno v. American Civil Liberties Union, 521 U.S. 844 (1997)(striking down anti-indecency provisions of the Communications Decency Act).

[70] See, e.g., New York Times Co. v. Sullivan, 376 U.S. 254 (1964) (establishing actual malice standard for libel actions by public figures).

[71] The importance of relationship-specific investments in international trade as received a fair amount of attention through the years. See, e.g., Beth V. Yarbrough & Robert M. Yarbrough, Cooperation and Governance in International Trade (Princeton University Press 1992); Pol Antras & Robert W. Staiger, Offshoring and the Role of Trade Agreements, 102 Am. Econ. Rev. 3140 (2012).

these risks into consideration and price them before investing. When the risk in question is *inefficient*, however, as I have assumed treaty violations to be, joint gains arise from eliminating the risk and averting its discouraging effects on investments.

These observations offer an additional perspective on the utility of as such challenges to measures that create a significant probability of a violation, particularly if the violation poses a significant threat to sunk investments. Consider ten measures that each create a 10% probability of a treaty violation that will substantially impair returns to sunk investments. Only one of these measures would be expected to meet with an as applied challenge, yet all of them may create significant inefficiencies if the *ex post* remedy following an as applied challenge does not afford full compensation. Once again, other things being equal, the utility of as such challenges will tend to increase as the probability of a violation associated with a measure increases, and as the magnitude of the harm associated with a treaty violation increases.

Concern for a chilling effect was explicit in the *U.S. – Section 301* case at the WTO. The panel made much of the fact that the possibility of a treaty violation created uncertainty in the marketplace and added to the risk faced by private transactors. "Such risk or threat, when real, was found to affect the relative competitive effect between imported and domestic products because it could, in and of itself, bring about a shift in consumption from imported to domestic products.[72]"

Recognition of this problem is a useful step forward, but by itself insufficient to afford a reliable guide for policy. Every potential treaty violation creates some marketplace risk. The key question is the magnitude of the problem, which turns on the factors developed above.

Conclusion

This paper offers some simple economic observations about the utility of as such or facial challenges. They suggest that the oft-cited "rules" of the WTO and U.S. Constitutional systems regarding the availability of as such challenges make little economic sense, while some

[72] U.S. -- Section 301, Para. 7.84.

recent decisions that depart from these rules have plausible economic justification. More careful attention to the relevant costs and benefits of as such challenges going forward may enable the decisional law to evolve further in an economically sensible direction.