Words Matter: How WTO Rulings Handle Controversy

Marc L. Busch†       Krzysztof J. Pelc‡

April, 2016

Abstract

Judicial rulings handed down by international courts have long attracted the attention of those looking to see “who wins?”, but much more is at stake. Like other institutions, the World Trade Organization (WTO) is widely seen as offering rulings that balance legal discipline against political constraints, yet the literature has offered scant direct evidence of this. We argue that one way in which the WTO handles politically sensitive issues is by increasing the amount of affect in their rulings. In doing so, judges provide national governments with the linguistic resources they need to persuade their domestic audiences of the legitimacy of compliance. To test our expectations, we conduct a text analysis of all rulings rendered by the institution since 1995. Specifically, we find that more politically charged decisions, such as the ones concerning non-fiscal rather than fiscal aspects of national treatment claims, are explained in qualitatively different terms. When faced with these politically sensitive issues, judges rely on judicial economy four times as much, on citations to past cases twice as much, and they employ significantly more affect-laden terms. We also find that as an issue gets ruled on repeatedly, the amount of affect deployed progressively goes down. In sum, the WTO chooses its words strategically to persuade litigants, and their domestic audiences, of the merit of compliance in politically fraught disputes.

†School of Foreign Service and Department of Government Georgetown University, mlb66@georgetown.edu
‡Department of Political Science, McGill University, kj.pelc@mcgill.ca
1. Introduction

When examining the rulings of international courts, most attention gets paid to the direction of the ruling: who won, and who lost. Yet this overlooks the considerable lengths to which courts go to explain their decisions. Courts seek to persuade as much as to judge—especially in international law, where the legitimacy of unelected individuals ruling on matters of national sovereignty is often called into question. International courts are thus at their weakest when ruling on matters that touch on sensitive domestic issues that can be seen as encroaching on sovereign authority; they have reason to fear pushback from national governments, in the form of noncompliance, or in efforts to reformulate the rules themselves.

Students of courts have shown that in dealing with such politically charged matters, courts behave strategically, by making greater use, for instance, of rhetorical sources, or references to prominent authors and texts.¹ They can rely to a greater extent on citations to past cases, vying for the added legitimacy that comes with the demonstration of consistency with an established body of law.² Courts can also curtail the scope of the ruling as a means of limiting its jurisprudential impact.³ In this article, we suggest that another means by which courts seek to persuade is by relying on affect, that is, words denoting overt positive or negative sentiment.

In doing so, we focus on a court which is rarely associated with sentiment: the World Trade Organization (WTO). Even as its task is to rule on technical economic issues, WTO panels and the Appellate Body (AB) often use affect-laden words in an effort to help give greater force to their rulings. Lacking enforcement power, the WTO must persuade governments, and especially their domestic constituents, of the merits of compliance. We argue that one of the reasons

---

¹ Hume 2006.
³ Busch and Pele 2010.
for the deployment of affect is to provide governments with discursive resources to push for the legitimacy of compliance.

In 2009, faced with the third consecutive dispute against the United States on the same technical issue over the zeroing of dumping margins, the WTO's Appellate Body's ruling struck out in exasperation: "there comes a time when it is more important for the system of dispute resolution to have a definitive outcome, than further to pick over the entrails of battles past." In instructing the US to stop picking over the entrails of battles past, the AB sought to persuade not merely by reference to agreed upon law, but rather through the invocation of strong sentiment. We seek to identify the circumstances under which WTO judges become most likely to rely on sentiment in this way.

We argue that the incentive to respond to political controversy through increased affect distinguishes the WTO from other courts. The US Supreme Court, for instance, is found to strategically render its sensitive opinions less legible to make their review by legislators more difficult. The comparative legal literature describes how ambiguity may be another response to political sensitivity, as a means of offering defendants more leeway in their response to the ruling, and hedging against outright noncompliance. European courts, in turn, have been argued to use law as a "mask" for politics, dissimulating the political charge of their opinions in legal garb.

The WTO faces somewhat different incentives. Rather than pushing its rulings past a legislature or national governments, so that lower courts may enforce them, the WTO seeks to persuade governments to abide by its recommendations. And governments, in turn, have to sell any resulting policy reform to their domestic audiences. Indeed, much of the WTO's design is based

---

5 See, e.g. Staton and Vanberg 2008.
around this form of collusion between governments and the institution, as a solution to domestic political challenges. States join international institutions to push through Pareto improving reform which they know to be unpopular with vested domestic interests. The fact that trade concessions are made in a reciprocal fashion; that negotiations during disputes are conducted behind closed doors; and that the final remedy in legal challenges involves the imposition of the very kind of barriers that the institution is created to do away with—all these features result from a recognition by the institution's designers that the WTO serves to strengthen the government's hand vis-à-vis its domestic import-competing industries. Keeping with this institutional design, our claim is that the WTO handles political controversy not through obfuscation, but by providing governments with the discursive resources they need to push for compliance domestically. Specifically, it couches its opinions with affect-laden reasoning that can be taken up by compliance-oriented constituencies within the defendant country. In this respect, our reasoning shares much with Lupu and Voeten (2012), who argue that "political actors are unlikely to be persuaded by legal justifications"; we would claim that political actors are likely to find non-legal content most useful in persuading their own audiences of the legitimacy of compliance.

The WTO employs greater affect alongside other strategies for dealing with political controversy already identified in existing studies. That is, rhetorical reliance on affect is not used to the exclusion of these other methods, and our empirical analysis aims to show them being employed side by side. Judges still have an incentive to buttress legal opinions that are likely to face push-back with greater reliance on formal citations to past rulings, but this strategy is unlikely to be addressed to the same audience as the deployment of affect. And both the deployment of affect and citations to past cases are used net of issue avoidance methods: WTO judges are known to rely on judicial economy as a means of reducing the number of findings of violation in politi-
cally charged disputes. Overall, our empirical strategy consists in showing that greater affect is used alongside these other known strategies, to deal with the same type of politically charged cases.

In doing so, we rely on a novel set of text analysis methods, which allows us to test inferences on the entire corpus of WTO law at once. Legal rulings are uniquely well suited for content analysis: law is a fundamentally rhetorical exercise, and legal rulings follow a consistent structure, making computer-assisted analysis easier. While automated text analysis is only beginning to be used to study the WTO, it has already been employed to shed light on the actions of other institutions where the choice of words looms large, such as central banks. In particular, Grimaldi (2011) shows that European Central Bank (ECB) bulletins can help to predict periods of financial stress, while sentiment analysis of the minutes of the Swedish Riksbank is useful in anticipating that bank’s subsequent policy decisions. Likewise, Sauter et al. (2013) argue that the ECB strategically reduces the number of words denoting uncertainty in the midst of market distress. These scholars reach a conclusion similar to the one we arrive at in this paper: social and political actors are strategic in the words they use. We argue that the same is true of the WTO, and international courts more generally.

We conduct the first automated sentiment analysis of all WTO rulings issued since 1995. We look to see whether affect terms, as identified in a sentiment dictionary used in the political communication literature, are more frequently cited in some dispute than in others. We demonstrate that rulings vary a great deal in the amount of sentiment they express, in ways that follow expectations. The more politically sensitive a legal issue, the more heavily judges rely on affect-laden words in their rulings. We show how health and safety standards disputes like *EC—*

---

7 Apel and Grimaldi 2012.
Hormones, for instance, which are arguably the most controversial in the WTO’s caseload, lead the way in the frequency of affect terms. In sum, the WTO is concerned not only about what it says, but how it says it.

To test our theory, we perform a series of sentiment analyses of WTO rulings. We show that areas traditionally viewed as more politically sensitive are associated with panel and AB rulings that rely more heavily on affect. Since these different legal issues also vary in a number of other ways that are difficult to control for, we then focus on disputes within a given issue area: national treatment. We divide all national treatment cases between fiscal and non-fiscal matters, corresponding to GATT III:2 and III:4, respectively. The latter are typically more politically charged because they involve less transparent regulatory issues that bear on national sovereignty. But both are legally axiomatic claims pertaining to the same legal principle: the “equality of competitive conditions” between foreign and domestic producers. This sets up an ideal test of whether the WTO strategically frames certain key rulings in more affect-laden terms than others. Finally, we are also interested in how legal opinions vary over time, as they repeatedly confront the same sensitive legal issue. Do courts progressively normalize these issues? One way of assessing this is to measure whether the affect deployed to deal with a politically sensitive issue differs through time, as the issue is repeatedly ruled on.

To preview our results, not only are SPS and TRIPs disputes, conventionally thought of as dealing with the most controversial issues in the trade regime, associated with greater affect, but we find similar variation within the single issue of national treatment. There, we find far greater affective content in panel and AB verdicts concerning non-fiscal, rather than fiscal, national treatment cases. Consistent with these results, WTO panels are much more likely to exercise judicial economy when they rule on non-fiscal measures. Judges also offer more citations to
past cases when explaining their rulings on non-fiscal measures. We also find consistent variation within legal issues, across time: considering all legal claims put forth to the WTO, we find that judges are more likely to deploy high levels of affect when first dealing with a new issue, and come to rely on affect progressively less as they repeatedly deal with the issue. The implication is that over time, the court may have a normalizing effect on the treatment of politically charged disputes. Taken together, our findings provide a fuller picture of how one international court handles politically sensitive issues.

**2.1 Affect, Persuasion and International Courts**

Words give meaning and context to states’ legal obligations. As Hume (2006) puts it, “though sometimes overlooked, the language of court opinions can be as important as the disposition of cases”. More generally, Young and Soroka, on whose methods we partly rely in our empirical analysis, contend that “[p]olitical discourse cannot be reduced to mere factual information—the tone of a text may be as influential as its substantive content.” Indeed, courts seeking to affect policy must communicate the substance of its ruling in a persuasive manner. This becomes especially difficult in the case of international courts who cannot rely on lower courts to implement their decisions. International law is based on consent; states must accept the short- and long-term implications of a decision. With this in mind, judges seeking to apply the law in a politically palatable way have an incentive to set the right tone for their opinions.

Litigants themselves also worry about the content of rulings, beyond their direction. This is especially apparent in the case of the WTO. For example, WTO complainants sometimes ap-

---

8 Hume 2006.
9 Young and Soroka 2012, 205.
10 Steinberg
peal their own victories, given concerns that the words used by the panel do not suit their long-term interests, as when the US challenged its own victory in *EC—Aircraft* because the panel did not see Airbus subsidies as a “programme.” Indeed, some cases get a lot of attention not because of the outcome *per se*, but because they modify the meaning of shared rules and set up the prosecution of broader “offensive” interests, as in *US—Tuna II*. In a similar vein, both sides typically have strong preferences over the panel’s use of judicial economy, even though this bears on the *scope* of the verdict, not its direction.

Students of the US Supreme Court have long paid attention to how the words used in legal opinions reflect judges’ strategic behavior. Increasingly, this work is turning to the tools that we rely on here. Owens et al. (2013) automatically code Supreme Court opinions according to readability and show that the level of obfuscation of cases is a function of Congress’ resistance to the outcome, as well as lower court conflict over the matter. Most recently, Corley and Wedeking (2014) use a dictionary of certainty to machine code levels of authoritativeness of US Supreme Court opinions. They find that opinions that rank higher in their degree of certainty are more likely to be followed by lower courts. These empirical findings are related to an earlier series of theoretical arguments, by authors like Staton and Vanberg (2008), who argued that courts may strategically rely on vagueness as a means of hedging against potential non-compliance. Ambiguity, they claimed, makes it more difficult to detect when lower courts fail to follow precedent.

These expectations contrast markedly, and in instructive ways, with our own. In the case of domestic courts with powers of judicial review, like the US Supreme Court, the expectation is most often that the court responds to political sensitivity through obfuscation: when anticipating

---

1 Pelc 2014.
12 Busch and Pelc 2010
political pushback, it renders more complex, less readable rulings containing more legalese. What these studies highlight is that the US Supreme Court speaks most directly to the lower courts that will be tasked with implementing their rulings, while trying as much as possible to avoid political opposition from legislators. The court's incentives in this respect are similar to those of the European Court of Justice, which speaks to members' national courts, while trying to avoid pushback from the national governments whose laws they are reviewing.¹³

The picture looks very different in the WTO, the goal of which is to reduce discriminatory barriers to trade, the demands for which comes from domestic groups. We expect that the rhetorical function of legal rulings at the WTO should be consistent with the design of the institution. Most trade scholarship views the WTO as a means for governments to increase their bargaining power in dealing with powerful industries at home who seek protection against import competition.¹⁴ This is why countries join international institutions and appoint third party enforcement mechanisms in the first place. Much of the WTO’s legal design, from the fact that pre-trial consultations are in camera,¹⁵ to the procedures for retaliating against a recalcitrant defendant’s most influential firms and industries,¹⁶ reflects a recognition by the institution that domestic politics are decisive. Specifically, governments seeking outcomes that favor the country as a whole use international commitments to push through reforms that may displease concentrated import-competing industries. In this way, studies of the WTO have claimed that governments may even welcome some findings of violation, if it affords them political cover to remove distortionary support of vested interests. The institution thus serves to tie the hands of governments

¹³ Garrett, Kelemen and Schultz 2002.
¹⁵ Busch and Reinhardt.
¹⁶ Bown.
vis-à-vis these groups, and allow political leaders to deny their demands for distortionary protection.

We argue that the rhetorical content of the WTO's rulings plays an analogous function: in politically-charged cases, the words used by the court provide discursive resources to governments to question the validity of a domestic policy. In the trade regime, the “audience” that ultimately requires persuasion is made up of domestic constituents. When speaking to their audience, governments themselves opt for affect-laden terms over legal citations: they invoke the authority of a court, and the content of its ruling, to delegitimize the measure at issue, and push for the legitimacy of compliance. The court and the defendant thus form an unlikely alliance: together, they must persuade domestic audiences of the legitimacy of compliance.

In an oft-quoted phrase, the WTO has neither truncheons nor tear gas.\(^\text{17}\) More pointedly, it also lacks lower courts, as in the case of the USSC, or national courts, as with the ECJ, to implement its rulings. For this reason, our expectations over the functions of judicial language differ from those of studies on the USSC. We expect the response to politically charged cases to be not obfuscation, but rather an attempt at persuasion. In such cases, WTO judicial bodies will seek to explain their decisions by deploying greater affect.

Consider a recent case, over the labeling of US meat products, that led to a great deal of political handwringing. In 2008, Canada accused the United States of having implemented overly restrictive meat labeling requirements, within the meaning of the Technical Barriers to Trade (TBT) Agreement. The US labeling requirements, called country-of origin labels (COOL), were presented domestically as a food safety measure informing consumers about where their meat was born, raised, and slaughtered. The Panel ruled, and the Appellate Body upheld (albeit modi-

\(^\text{17}\) Bello 1996.
fying its reasoning), that the US labeling measure was inconsistent with TBT Article 2.1, because it accorded less favourable treatment to imported livestock than to like domestic livestock.

The resulting ruling ranks among the highest for the degree of affect in our dataset. Indeed, in formulating its ruling, the AB insisted on how the US measures imposed a “disproportionate burden,” a “record keeping burden,” and an “undue burden”, all of which led to the measure's “detrimental impact.” The information provided by the labels was "arbitrary, and the disproportionate burden imposed ... unjustifiable." The choice of these terms fell within the court's discretion, these terms having have no textual support within the GATT texts, or the TBT Agreement.

The US, as all governments on the losing end of a dispute, claimed to be disappointed with the outcome of the dispute. So did several domestic interest groups, though some conceded that the ruling suggested the need for policy change. As the American Farm Bureau Federation put it, "While we had hoped the WTO would rule favorably on the regulatory changes made to COOL in recent years, it is now clear that is not the outcome–and it is time to bring this challenge to a final resolution." 19

What is relevant to our argument is what happened next. In arguing for the repeal of the policy at issue, legislators in the House of Representatives invoked the WTO ruling repeatedly, and repeatedly drew on the affect-laden language of its opinion. Michael Conaway, a Republican representative from Texas, argued on the House floor that "The Appellate Body agreed with the panel that the recordkeeping and verification requirements of the amended COOL measure im-

19 American Farm Bureau Federation. May 27, 2015. Letter to the Chairman of the House Committee on Agriculture.
pose a disproportionate burden", that it had "a detrimental impact", that it imposed a "record-keeping burden", which in turn had additional "detrimental impact".\footnote{Opening Statement. Michael Conaway (R-Texas). H.R. 2393. All emphasis our own. \url{https://agriculture.house.gov/opening-statements/opening-statement-chairman-k-michael-conaway-hr-2393-%E2%80%93-repeal-country-origin} Last accessed June 8 2015.} Another legislator, Jim Costa, a Democratic Representative from California, chimed in, claiming that "the World Trade Organization’s experience to demonstrate that COOL is detrimental to our state and national economies".\footnote{http://www.cattlenetwork.com/advice-and-tips/cowcalf-producer/house-ag-committee-wastes-no-time-act-cool Last accessed June 8 2015. Emphasis our own.}

Our point is that the WTO's deployment of affect allowed US legislators to press the case that "Repealing COOL is the right thing to do," as David Rouzer, the Representative from North Carolina, put it. This was not just about the WTO's finding of noncompliance with the TBT Agreement, it was about demonstrating the illegitimacy of the policy to non-legal audiences. And to be sure, these statements by legislators, which were drawn from the WTO's discretionary use of high affect terms with no textual support in WTO texts, were also then in turn picked up by media aimed at the agricultural community\footnote{See for instance, Pork Network, "House Bill Would Repeal COOL Requirements; NFU Continues Defense", which transcribed Rep. Conoway's statement, above. \url{http://www.porknetwork.com/news/house-bill-would-repeal-cool-requirements-nfu-continues-defense} (Last accessed April 10. 2016.)}—the segment of the domestic audience most at issue.

The domestic discourse in the wake of the COOL dispute illustrates the unlikely alliance that binds the WTO to defendant governments, or to pro-compliance constituencies within those governments. Much of the WTO's design aims to strengthen this alliance. Retaliation lists, to name but one institutional feature that also played an important role in the COOL dispute, appear as an unlikely part of the WTO architecture: after all, the WTO is in the business of abating trade barriers, and retaliation brings them up. But retaliation lists harness domestic politics in the serv-
ice of liberalization, by enlisting the support of export-oriented interests that fear being hit by retaliatory duties. Governments then make a show of having their hands tied by the threat of retaliation, and argue for the need to comply to avoid hurting exporters. Affect plays an analogous function, by providing governments with discursive resources to draw on in arguing for compliance.

The US debate over COOL is thus a striking example of the judicialization of international relations in practice. US Republican legislators—far from traditional supporters of the WTO, or international courts more generally—are seen quoting directly from a WTO ruling to compel domestic policy change in a sensitive regulatory area. In doing so, they rely not on the legalistic aspects of the opinion, according to which the US measure accords less favorable treatment to imported livestock; instead, they pick up on the most affect-laden parts of the report.

To be sure, not everyone in the US was so quick to call on reform in deference to the WTO ruling in US—COOL. But tellingly, those who opposed compliance did not cite the ruling's affect-laden language; they made no reference to "detrimental impacts" or "disproportional burdens". They stuck to the legal facts and legal terminology. As the Ranchers-Cattlemen Action Legal Fund (who were opposed to swift compliance before retaliation amounts were set) characterized the loss, the WTO "found Canadian and Mexican livestock were treated less favorably than domestic livestock."\(^{23}\)

How common is it for courts to rely on affect in this manner, in an effort to provide pro-compliance constituencies within defendant governments with the linguistic resources to push for the legitimacy of compliance? This is what we attempt to measure in the empirical analysis.

2.2 What Are the Costs of Affect?

We have argued that greater reliance on affect increases the persuasive power of rulings, in that affect conveys to domestic audiences not only the non-legality, but also the illegitimacy of a sensitive trade measure. This naturally leads to the question, if it is such an effective rhetorical device, why would courts not always rely on maximum affect?

Courts face competing incentives between seeking to secure compliance in the case at hand, and seeking to influence the direction of case law. The cost of deploying affect arises from this tradeoff. Specifically, the type of rhetorical tools that judges deploy to push for compliance—in this case, through greater reliance on affect—also render their rulings less generalizable beyond the dispute at hand. To use our initial Zeroing example, the AB may have finally swayed US legislators by instructing them to stop "pick[ing] over the entrails of battles past", but the choice of language made it unlikely that this invective would be picked up in subsequent rulings on related legal questions. This aspect of the ruling was aimed at resolving the disagreement at hand and at surmounting US recalcitrance, rather than directing jurisprudence on, e.g., the interpretation of the AD Agreement. In this isolated case, the AB's rhetorical flourish may well have paid off. Following this third dispute over zeroing, the US did eventually comply with the sum total of findings against it. Yet that affect-laden passage is unlikely to travel beyond the case in question. Affect is context-specific.

The choice judges make over the terms they use in their opinions mimics the greater tradeoff courts face in weighing the present case against its future implications on jurisprudence. Most often, judges fear that a ruling may be seen by states as exerting too great a constraining effect over the shared understanding of the rules. For example, in determining whether two goods are “like” for the purposes of national treatment, and thus whether any favorable treatment
to one must be extended to the other, the WTO explains that the “... concept of ‘likeness’ is a relative one that evokes the image of an accordion. The accordion of ‘likeness’ stretches and squeezes in different places as different provisions of the WTO Agreement are applied.”

In this sense, the wider the accordion is stretched, the more robust the implications for follow-on litigation, a point the panel in Japan—Alcoholic Beverages II, for instance, went to pains to explain was not its intention.

Likewise, in EC—Hormones, the panel went out of its way to limit the repercussions of its ruling, demanding that, “[i]n order to avoid any misunderstanding as to the scope of implications of the findings above, we would like to stress that it was not our task to examine generally the desirability or necessity of the EC Council Directives in dispute.” Quite the opposite, the panel insisted on how it was limited by those aspects of the European measures raised by Canada, and was “further limited to the specific provisions of GATT and the SPS Agreement which have been invoked by the European Communities in support of this import ban,” the point being that other countries that ban, or use, these hormones should not read too much into the decision. Along these same lines, the panel explicitly attempted to carve out its ruling from broader jurisprudence over domestic regulation, insisting that "The ability of any Member to take sanitary measures which do not affect international trade was not at issue in the present case."  

This tradeoff between securing compliance in the case at hand and influencing subsequent jurisprudence also governs the deployment of affect. Yet here, affect is a case-specific investment in politics. By deploying greater affect, the court addresses itself to non-legal audiences over the case at hand, and thus concedes some degree of influence over how legal actors

---

25 ibid.
handle subsequent related cases. It aims to persuade states of the illegitimacy of a policy in a politically sensitive context, but it does so at the cost of generalizability.

2.3 Parallel Judicial Tools

In this article, we focus on the role that affect play in judicial opinions. Yet this is not to deny the importance of other parallel means by which judges handle political sensitivity when ruling against sovereign states. How do these different methods relate to one another?

Prior to formulating an opinion, judges can also engage in issue-avoidance. Scholars have argued that the ECJ restrains its jurisprudence when a ruling might otherwise run afoul of key members’ sensibilities.27 The WTO's AB, too, is argued to do this, striking a balance between issuing legally consistent rulings, and being flexible enough to secure compliance.28 In fact, issue avoidance happens before a ruling is issued. In nearly half of all WTO disputes, panels exercise what is called judicial economy.29 Under the guise of saving deliberation time and judicial resources, panelists can choose not to rule on a given claim, if they can do so while still securing a “positive solution” to the dispute at hand. Specifically, if the losing party is already found to be in violation of one claim, panelists can invoke judicial economy to skip a finding of violation on an associated claim. This is political; panels use judicial economy to narrow the scope of a verdict that would otherwise be perceived as controversial by the wider Membership. We take this as our departure point; after a court engages in issue-avoidance, it still has to explain itself in the rulings it issues. This is where affective framing comes in.

Another means at judges’ disposal is reliance on citations to past cases. Here too, analo-

27 Garrett, Kelemen and Schulz 1998, 150.
gies can be drawn to the ECJ: work on what has been called the European constitution court has shown that judges are more likely to rule against powerful member government if they can rely on existing precedent when doing so. Larsson et al. (2015) also show that strategically increase the number of prior rulings they rely on when handing down rulings that are especially likely to be politically sensitive. Similarly, we expect judges to couch their opinions in citations to past cases when ruling on sensitive issues of national sovereignty, vying for the added legitimacy that comes with the demonstration of consistency with an established body of law. Here, the objective is as much to convey the lack of agency over a decision as to demonstrate its inevitability.

We expect these methods to be used in conjunction with one another. Where we see more affect-laden language, we also expect more use of judicial economy, and greater reliance on citations to past cases. The objective of all these methods is the same: to balance the competing pressures between legal obligation and political constraint in disputes that implicate national sovereignty.

3.1 Measuring Affect in the WTO

What does affect in legal rulings look like, and how to we measure it? We begin by offering a picture of affect across WTO rulings, set out by legal issue. To do this, we rely on a “sentiment dictionary” created by Soroka and Young, about which we say more in our empirical section. We use this dictionary to calculate the number of positive and negative sentiment words in every ruling. We scale these scores by the total number of words, to obtain a relative frequency of affect. Negative words that come up in WTO rulings have roots such as “inconsist”, “injur”, “violat”, “restrict”, “prohibit”, “against”, “risk”, “contrary”, “discrim”. Positive words have roots such as

---

30 Garrett, Kelemen and Schultz 1998.
“agree”, “respect”, “relevant”, “benefit”, “adopt”, “rights”, “principl”. We sum the two to obtain a relative frequency of total affect. Figure 1 shows the distribution of affect across WTO rulings, sorted by legal issue. It also distinguishes between panel and AB rulings. Legal issues at the top have the greatest frequency of affect-laden words, and legal issues at the bottom have the least.

The first thing to note is that there is significant variation in the frequency of affect across legal issues, even as each of these represents a great many (and often heterogeneous) disputes. The average dispute over services has 25% less affect than the average health and safety dispute, the latter of which fall under the Sanitary and Phyto-Sanitary (SPS) Agreement. Both legal issues correspond to more than 20 disputes each, and the difference in affect between them is highly statistically significant. That this particular legal issue ranks highest in terms of affect across all WTO legal issues makes sense. SPS disputes concern health and safety standards, and are likely among the most controversial disputes at the WTO. Health and safety standards are quickly becoming the protectionist barrier of the 21st century, especially among developed countries. The notorious EC—Hormones case is indicative of this.

The second highest-ranking issue, in terms of affect, concerns intellectual property issues under the Agreement on Trade-Related Aspects of Intellectual Property Rights, or TRIPs. Here too, judges face matters of considerable political sensitivity, bearing on a Member’s policies on generic drugs, for example, its laws on patentability, and the need for its courts to enforce deterrent-level sanctions. The other striking thing to note in Figure 1 is that while AB affect co-varies with panel affect to a great degree (the bivariate correlation is 0.60), AB affect tends to be significantly higher. This will come as a surprise to students of the WTO who are used to thinking that the AB is the more staid, coherent legal body in the WTO, given that its judges have tenure, and its rulings are final. We submit that it is precisely because of these factors that AB judges
deploy more affect than *ad hoc* panelists.

**Figure 1: Average Affect Across Agreements**

![Figure 1: Average Affect Across Agreements](image)

These measures remain largely illustrative. Indeed, the broad legal issues listed in Figure 1 are different from one another in a number of ways, complicating direct comparisons. Some concern more recent agreements, while others may be thought to be more technically complex. To conduct a more controlled analysis, we focus on a single legal issue: the question of *national treatment*, which is taken up by GATT Article III. National treatment is one of the two axioms of non-discrimination in WTO law, the other being most-favored nation (MFN), which corresponds to Article I GATT. MFN prevents differential treatment *across* Members, insisting that, in terms of “any advantage, favour privilege or immunity” granted by one Member to another, it must be granted to all Members. National treatment is about the “equality of competitive conditions” between foreign and domestic products *within* a Member’s economy. The architects of GATT pre-
dicted that, without national treatment, Members would simply replace lost tariffs with fiscal or non-fiscal measures. National treatment was thus meant to be an “anti-circumvention device,”\textsuperscript{31} inasmuch as it prevented Members from back-filling “at the border” protectionism using “within border” measures.

As in Figure 1, GATT issues fall around the median frequency of affect. Yet our contention is that we should observe the same variation within GATT III cases as we observe across all legal issues. This is because GATT III splits into two main provisions: GATT III:2 on fiscal and GATT III:4 on non-fiscal measures. The architects of GATT were mainly concerned that fiscal measures—i.e., internal taxes and other charges—would be used as substitutes for tariff reductions realized through successive trade rounds.\textsuperscript{32} Since the logic of GATT, writ large, was to cap and cut MFN tariffs, convert all non-tariff barriers to MFN tariffs and cut these too,\textsuperscript{33} the burden fell on national treatment to ensure the value of these tariff bindings. GATT III:2, \textit{first sentence}, covers “like goods,” for which differential taxation is, on its face, protectionist. GATT III:2, \textit{second sentence}, covers products that are “directly competitive or substitutable,” for which differential taxation above a \textit{de minimus} threshold, and applied so as to afford protection, is illegal.

The architects of GATT also worried about non-fiscal measures, including “all laws, regulations and requirements” bearing on commerce. These measures were less common at the time, and isolating their trade effects was difficult because they generally do not work through prices. This meant that GATT III:4 would have to deal with measures that often look like what, under GATT Article XX, is called a “disguised restriction” on trade. Subsequent WTO texts, notably on Technical Barriers to Trade (TBT), which traces to the 1970s, helped catalogue these different

\textsuperscript{31} Zhou 2012
\textsuperscript{32} Mavroidis
\textsuperscript{33} Mavroidis
types of non-fiscal measures. The common denominator, though, is that non-fiscal measures include regulatory policies that may not be thought to be trade-related, never mind a disguised restriction on trade. For our argument, the point is that non-fiscal measures are typically more politically charged than fiscal ones.

To be clear, we expect that panels and the AB will engage in at least some affective framing in all cases. But as with judicial economy, there is a cost to adding affect to a ruling. Sentiment is a discursive resource that pro-compliance constituencies, especially within the defendant country, may pick up on. But affect, as context, qualifies the substance of the ruling, thus limiting its contribution to WTO jurisprudence. In this way, affective framing involves the very same balancing of “flexibility” and “rigidity” as judicial economy, or the use of escape clauses more generally.\textsuperscript{34} We thus expect WTO judicial bodies to economize on emotive content, reserving greater investments of this sort for politically charged disputes.

This reasoning allows us to frame expectations about affective content in GATT III:2 versus GATT III:4 disputes. Given the more politically charged nature of the latter, we expect that GATT III:4 disputes will be associated with more reliance on affect-laden terms, net the greater use of judicial economy and stronger reliance on citations to past cases.

We also seek to account for the deployment of affect in issues beyond national treatment. Yet direct comparisons between different agreements are problematic. The concern, as above, is that some agreements may be "hardwired" for affect. To address this issue, we consider variation within each agreement, and across time. Specifically, we test the relationship between the legal novelty of rulings, and the affect deployed. The first time a given legal claim is brought by a complainant, the associated legal opinion looms especially large. The legal issue is untested, and

\textsuperscript{34} Rosendorff 2005.
there is greater ambiguity over the likely outcome of the ruling, which may come as unexpected to the litigants. The defendant may be less prepared for a finding of violation than when the same claim has been ruled on elsewhere. Conversely, as the claim gets ruled on repeatedly, defendants learn what to expect; litigants converge towards a shared understanding of the underlying claim. These factors contribute to increasing the odds of pushback against a finding of violation over a novel legal claim. Along these lines, Garrett, Kelemen and Schultz (2002) find that the European Court of Justice is less likely to find against powerful governments when it is short of precedents to rely on. A novel claim constitutes the absence of such precedent; when ruling on these, judges do not have the political that comes from prior rulings.

For this reason, rulings over novel legal claims, all else equal, are also more likely to prove decisive, and influence subsequent interpretations of the same clause. In this respect, what we describe above as a cost of relying on affect—reduced relevance for jurisprudence—may be seen as a benefit: by concentrating on resolving the case at hand, and tempering its jurisprudential impact, the court may eschew some of the sensitivity that comes with an early ruling on a given claim.

Taken together, these considerations lead us to expect that more novel legal claims should lead to more affect-laden opinions, both as a means of pushing for compliance in the face of political salience, and as a means of managing the jurisprudential impact of the case. The implication is that judicial rulings have a normalizing effect over time. Through their rulings, courts reify expectations and lay down norms of behavior. Initially, controversial issues require a great deal of persuasion on the part of the court, which we proxy for through the amount of affect deployed. As the court keeps ruling on a given issue, actors’ expectations converge and become more firmly set, and the need for persuasion is lessened. The question is whether we see an asso-
ciated drop in the amount of affect deployed over time. The nth time the WTO adjudicates an issue that was at first politically charged, does it become business-as-usual, with no need for affect-laden language?

3.2 Empirical Analysis

To test our expectations, we perform automated sentiment analyses of every WTO ruling, both by panels and by the AB, from 1995 to 2014. We code the amount of positive and negative sentiment across the entire body of WTO jurisprudence. After illustrating the overall picture of affect across different legal issues in Figure 1, we then confine our analysis to fiscal versus non-fiscal national treatment cases. Finally, we examine how the amount of affect in legal rulings varies over time, within a given legal issue.

Throughout, we rely on an established sentiment dictionary through Lexicoder, which classifies texts according to sentiment in a way that has been shown to be highly consistent with human coders, improving significantly on prior dictionary methods. The stated benefit of Lexicoder is its compatibility with a range of different discourses. As Young and Soroka explain, Lexicoder provides “a comprehensive valence dictionary,” the objective of which is to be “broadly applicable for scholarship on the tone of political communication.” About 2300 words within the Lexicoder sentiment dictionary are flagged at least once in the WTO texts.

Importantly, we confine our analysis to the judicial reasoning in every panel or AB report. This is important, since panel reports, even more so than AB ones, contain a number of voices, as both litigants make their case and third parties chime in to offer opinions. Since we are strictly interested in judges’ language, we manually extract panelists’ and AB judges’ words from each

35 Young and Soroka 2012.
ruling. Given the dialogue-like structure of these reports, we preferred not to rely on machine coding for greater accuracy. The result is a corpus of 8,101,996 words, of which 6,342,876 come from panelists and the rest from AB judges. This underscores the paper’s premise: the WTO has offered a lot of content in explaining its decisions, yet scholars have largely confined themselves to looking at which country won or lost in a given dispute.

Looking at the disputes at both extremes of the spectrum of affect is telling. On one end, the three least affect-laden disputes are EC—Customs Classification of Frozen Boneless Chicken Cuts; EC—Selected Customs Matters; and European Communities—Tariff Treatment of Certain Information Technology Products. All three involve well-established rules over countries’ schedules of concessions, and could not be less controversial. By contrast, the three highest affect cases are US—Softwood Lumber, Mexico—High-Fructose Corn Syrup (HFCS), and US—Byrd Amendment. All three are notorious disputes that dealt with politically sensitive markets, as in the case of softwood and HFCS, or sensitive domestic policy, as in the case of the Byrd Amendment. All three also took on sensitive legal issues that had not been frequently ruled on previously.

In fact, when we count the average number of prior rulings that have been rendered on the legal claim at issue in each set, we see a striking pattern: the three lowest affect cases build on an average of 8.1 prior rulings. Compare this to the 1.3 times the average claim raised has been ruled on in the set of highest affect cases. In other words, not only are these high affect opinions related to what are conventionally thought of as controversial legal issues, but they also appears to be associated with more unprecedented claims. Of course, one needs to verify that factors like time and dispute outcomes are not driving this relationship, and see whether it applies to the caseload as a whole. We do this below.
We start by summing the total frequency of positive valence and negative valence words to obtain a total frequency of affect. These co-vary: when WTO judges use more positive words, they also tend to use more negative words. Our argument applies to overall use of affect-laden terms in general, rather than either positive or negative sentiment. To be sure, the direction of our findings remains the same if we confine our analysis to either one or the other. But tellingly, the model fit is significantly better when we look at total affect. We confine our analysis to cases that are either GATT III:2 or GATT III:4 cases, to allow for a direct comparison between them. We code a *GATT III:4* indicator variable as being 1 for all GATT II:4 disputes, and 0 otherwise.

Our main findings are found in Table 1. Here we examine the relationship between the national treatment provision at issue in the dispute and the amount of affect deployed. Columns (1) and (3) examine affect in panel reports, while columns (2) and (4) consider AB reports. Since not all cases are appealed, the latter result in a smaller sample. Both samples are small, which is why we keep these models as parsimonious as possible. If anything, the size of the sample makes the significance of the findings all the more striking. Indeed, looking at the bare-bones univariate regressions in the first two columns, non-fiscal issues appear to result in significantly more affect than fiscal ones, as indicated by the positive coefficient on *GATT III:4*.

We then add two basic controls to both models. The first is a control for the year in which the dispute began, which could indicate a trend in affect over time. The second, *Pro Complainant Ruling*, is the proportion of claims in the dispute ruled in favor of the complainant, and against the defendant. This measure, which varies from 0 to 1, is calculated as the number of claims in favor of the complainant over the total number of claims decided. The measure reflects the well-established pro-complainant bias in the WTO: the average of this variable is 71% for panel rulings, and 66% for AB rulings.
The inclusion of these controls leaves our main result unchanged: non-fiscal disputes lead judges to rely significantly more on affect. According to this estimation, GATT III:4 cases produce 15% more affect in panel reports, and 19% more affect in AB reports, on average. The influence of time is negative for both panels and the AB, suggesting a downward trend in reliance on affect over time, but this appears to be significant for panels only. Interestingly, when the AB rules against the defendant, it relies on considerably more affect: AB judges employ an average of 40% more affect-laden terms when siding completely with the complainant than when siding completely with the defendant. The equivalent coefficient for panels shows no significant effect.

Figure 2: Citation Rates Within WTO National Treatment Claims

Next, we examine another of our theory’s testable implications. If rulings over non-fiscal issues lead judges to make greater efforts to persuade their audience, then this should also register on
reliance on citations. Indeed, while WTO law, like all of public international law, does not formally uphold the principle of binding precedent, judges routinely invoke past rulings to support their current findings. Following Hume (2006), in the case of the US Supreme Court, even when the sources being cited have no formal binding force, justices tend to deploy them to a greater degree when they are faced with “hard cases”, such as cases that overturn a statute. When faced with matters that tread on countries’ sovereignty, judges should be more likely to seek political cover by relying on citations to past cases. This is exactly what we find. Looking at Figure 2, WTO panelists cite more than twice as many cases, on average, in GATT III:4 cases than in GATT III:2 cases. This difference is statistically significant at the 1% level. Insofar as citations establish greater legitimacy for a legal opinion, judges appear to be seeking such enhanced legitimacy significantly more when ruling in politically charged cases.

Figure 3: Predicted Odds of Judicial Economy Within National Treatment Claims
Affect and citations are means that political savvy judges use to strategically shape the rulings they deliver. Yet prior to this, courts may seek to avoid ruling on a claim altogether. Our expectation is that GATT III:4 cases will also lead to greater rates of judicial economy. To test this, we create a Judicial Economy variable, coded as 1 if a panelist invokes judicial economy to refrain from ruling on any claim within the dispute. The data support our expectations. Figure 3 shows how the average rate of judicial economy is over 4 times higher in GATT III:4 cases. We also examine these odds in a probit model, shown in Table 2. Again keeping in mind the small size of our sample, the first column shows a univariate model. In the second column, we also include the proportion of the subsequent ruling found in favor of the complainant, since one might think that judicial economy is a means of avoiding additional opprobrium against a defendant already found in violation. And we include a variable of the defendant’s market size and income, since recent work on judicial economy has suggested that judges may be politically motivated in limiting findings against powerful defendants. Across both models, judicial economy is significantly more likely in non-fiscal cases. If we keep all other variables at their sample mean, then GATT III:4 cases are nearly 5 times more likely to invoke judicial economy to narrow the reach of their rulings. By comparison, neither market size nor wealth of the defendant appears to have any effect on judges’ propensity to avoid ruling on a claim.
Finally, we examine the evolution of the court’s reliance on affect through time, within a given legal issue. To do this, we code every legal issue addressed in every WTO dispute, at both the panel and the AB level. For each claim, we code the number of findings, and the direction of each finding. This allows us to say, for instance, that when China challenged the US poultry regime under SPS Article 8 in 2009, the court had no prior ruling to rely on—here was an entirely new legal issue. By comparison, when the US took on Chinese antidumping and countervailing duties in China—Autos on 2012, the WTO had produced over 20 findings across 8 different disputes under the relevant Article 3.1 of the Agreement on Antidumping. At this point, the scope of AAD Article 3.1 would be well known, and actors’ expectations would have adjusted. As discussed above, looking at the highest (lowest) affect cases in the caseload, these appeared to be over highly novel (established) legal issues. The question remains whether this is true on average: is a legal claim’s novelty associated with the amount of affect in the associated verdict?

Judging from the findings in Table 3, the answer appears to be yes. When we look at the relative amount of affect within a given provision, the number of times the claim has been brought up in the past appears significantly negatively related. As shown in Table 3, the more panelists and the AB rule on a given legal issue, the less affect-laden their reasoning becomes. This effect stands when we control for the direction of the ruling on that specific claim. The results also hold when we control for trends in time with cubic splines and when we add fixed effects on the legal agreement at issue, in columns 3 and 4 of Table 3.

In sum, we obtain a coherent picture of judicial behavior in the face of political controversy. In disputes bearing on national sovereignty, judges attempt to limit the reach of their rulings. On those claims they do rule on, they deploy greater affect, and greater reliance on citations.
to past cases, in an attempt to persuade litigant governments and their domestic audiences of the legitimacy of their findings. Finally, even on controversial claims, the more a given matter is ruled on, the less affect the court deploys. Following our theory, as a legal issue becomes progressively more established, judges rely progressively less on affect-laden words.

4. Conclusion

International courts must walk a fine line, hewing to the word of the law while being mindful of the political sensibilities of sovereign states. The resulting tension is exacerbated in the case of politically charged issues like health and safety standards, intellectual property protection, or domestic regulation. Scholars have identified a number of ways by which courts seek to manage this tension. Judges can seek to avoid making certain rulings through judicial economy. They can also defer to prior rulings by increasing citations to case law. We offer support for these beliefs in the case of the WTO, but we also argue that alongside these established judicial strategies, judges in the WTO handle political controversy by deploying greater amounts of affect, or non-legal terms denoting overt positive or negative sentiment.

We argue that this is an example of the way in which the court and the defendant form an unlikely alliance in the international trade regime: together, they must persuade domestic audiences of the legitimacy of compliance. This distinguishes the WTO from other legal systems, like the United States Supreme Court, or the European Court of Justice, where, by virtue of the existence of lower courts, or national courts, in charge of implementing rulings, legal opinions can address themselves principally to these other judicial actors. As a result, where scholars of the USSC and the ECJ expect legal obfuscation and legalese in the face of political controversy, we expect its opposite: reliance on affect-laden language, which can serve as a discursive re-
source for governments pushing for compliance.