Words Matter: How WTO Rulings Handle Controversy

Marc L. Busch    Krzysztof J. Pelc

Abstract

The rulings of internationals courts are often reduced to “who won?”, but much more is at stake. Like other institutions, the World Trade Organization (WTO) offers rulings that balance legal discipline against political constraints. 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 discursive resources 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. We also find that as an issue gets ruled on repeatedly, the amount of affect deployed progressively decreases. In sum, the WTO chooses its words strategically to persuade litigants, and their domestic audiences, of the legitimacy of compliance in politically fraught disputes.

Acknowledgements: We thank the editors, as well as participants at seminars at Northwestern University, the Lorentz Center in Leiden, and McGill University, for most useful comments. All remaining errors are our own.
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.¹

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.² 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.³ Courts can also curtail the scope of the ruling as a means of limiting its jurisprudential impact.⁴ 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 this way, we contribute to a better understanding of courts' calculations within what this Special Section calls "adjudication politics": the legal phase of judicialization.

¹ Garrett Kelemen and Schultz 2002.
² Hume 2006.
⁴ Busch and Pele 2010.
To do 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 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."\(^5\) 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.\(^6\) European courts, in turn, have been argued to use law


\(^6\) See, e.g. Staton and Vanberg 2008.
as a "mask" for politics, dissimulating the political charge of their opinions in legal garb.\footnote{Burley and Mattli 1993, Alter 2009.}

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 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.\footnote{Hudec, 1987. Mansfield, Milner and Rosendorff 2002, Maggi and Rodriguez-Clare, 1998, Staiger and Tabellini, 1999, 1987.} 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 by providing governments with the discursive resources they need to push for compliance domestically. 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 affect-laden content most useful in persuading their own audiences of the legitimacy of compliance.

Convincing these popular audiences of the validity of a ruling is likely to prove most difficult in cases that deal with areas of traditional sovereign control. These typically include behind-the-border issues, which comprise health and safety standards (e.g. hormones in meat), trade
related investment measures (e.g. domestic content requirements in media), government procurement (e.g. buy national campaigns), or environmental regulation (e.g. dolphin-safe tuna labeling). Since these touch on sensitive matters, they are also often the target of domestic groups looking for new avenues for import relief. Such "regulatory protectionism" has become one of the major concerns of WTO adjudication, as well as newly negotiated trade agreements. We expect WTO panelists and AB members to be sensitive to the domestic political difficulties that such cases present for governments, and calibrate the wording of their rulings accordingly. Rather than limiting themselves to legal findings of noncompliance, judges should offer justifications worded in affect-laden terms.

We test our beliefs by turning to automated text analysis methods, which allow us to test inferences on the entire corpus of WTO law at once. Legal rulings are uniquely well suited for automated content analysis: law is a fundamentally rhetorical exercise, and legal rulings follow a consistent structure, making computer-assisted analysis easier. 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 disputes 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 also show how health and safety standards disputes like EC—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.

---

9 Kono 2006.
10 See the Comprehensive Economic Trade Agreement between Canada and the EU, and the Trans-Pacific Partnership agreement between eleven nations.
We then 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.

We are also interested in how legal opinions vary over time, as judges 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 evolves through time, as an 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 significantly more affect being deployed in panel and AB opinions concerning non-fiscal, rather than fiscal, national treatment cases. This finding holds even once we control for the sentiment scores of complainant submissions, further reassuring us that the variation we have identified is not reducible to case type, but is instead being driven by judicial behavior. 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 a novel way by which 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”.11 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.”12 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.13 In focusing on the discretion that judges have over the wording of their opinions, and the impact this decision has on subsequent phases of adjudication, we pursue a set of questions related to Staton and Romero's contribution to this Special Section. Staton and Romero measure the varying clarity of legal opinions in the Inter-American Human Rights (IAHR) Court over remedies, and discuss the tradeoff that judges face in being more or less vague in their instructions over how to bring a policy into compliance.

---

11 Hume 2006.
12 Young and Soroka 2012, 205.
13 Steinberg 2004.
They find that vagueness is associated with lower rates of compliance, but that it may also be the result of judges offering countries more discretion in an attempt to leverage local knowledge over how best to achieve compliance. We identify a similar tradeoff in the deployment of affect by WTO judges. As we detail below, higher affect can be used to persuade domestic audiences of the legitimacy of compliance, but it may render rulings less generalizable beyond the dispute at hand, and leave the court open to accusations that it is straying from its mandate. In this way, variation in affect, much as variation in the clarity of legal opinions in the case of the IAHR, is driven by an attempt by courts to maximize competing considerations.

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 appeal 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

---

14 Pelc 2014.
15 Busch and Pelc 2010
(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 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. In all cases, courts anticipate how their rulings will be received and by whom, and they adjust their wording. As the audiences for rulings vary, so do courts' discursive means of anticipating pushback.

The picture looks very different in the WTO, where the goal is to reduce discriminatory barriers to trade in a way that is beneficial to the country as a whole, but politically difficult to achieve unilaterally. International trade institutions, and the third party enforcement mechanisms within them, are designed to provide a solution to this political challenge. Indeed, a dominant narrative in trade scholarship—from legal scholars like Hudec (1987), to economists like Staiger

\[\text{Garrett, Kelemen and Schultz 2002.}\]
and Tabellini (1987), and political scientists like Mansfield, Milner and Rosendorff (2002)—portrays the trade regime as a device employed by governments to credibly commit to behaving in accordance with the median voter's interests, despite powerful domestic interest groups who vie for trade protection.\textsuperscript{17}

Much of the WTO’s legal design, from the fact that pre-trial consultations are \textit{in camera},\textsuperscript{18} to the procedures for retaliating against a recalcitrant defendant’s most influential firms and industries,\textsuperscript{19} thus reflects a recognition by the institution that domestic politics need to be taken into account if countries are to achieve progressive liberalization. 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 findings of violation, if it affords them political cover to remove distortionary support of vested interests (Davis 2012). In other words, the design of international trade institutions recognizes the centrality of domestic audiences. We expect the rhetorical function of legal rulings at the WTO to be consistent with this design. WTO rulings are written not only with governments in mind, but also with a view to how those governments will persuade their own domestic constituents of the legitimacy of compliance.

The rhetorical content of the WTO's rulings becomes key in politically sensitive cases. These are cases that take on areas of traditional sovereign control, such as environmental or health and safety measures, or oppose an especially powerful domestic lobby, such as agricultural lobbies in developed countries. In an increasingly common twist, the two are often related: import-competing industries will turn to behind-the-border regulation that can offer them an edge against

\textsuperscript{17} See also e.g., Maggi and Rodriguez 2002
\textsuperscript{18} Busch and Reinhartd 2000.
\textsuperscript{19} Bown 2009.
foreign competition, and that is politically difficult to dismantle.\textsuperscript{20} Trade courts are increasingly called upon to adjudicate over these measures, by separating valid regulation arising from consumer demands from regulatory protectionism; in doing so, courts find themselves in the difficult position of ruling on governmental regulation. According to trade theory, pro-compliance actors within governments may be willing to amend regulation, but may be prone to domestic pushback in doing so.

In such politically charged cases, our claim is that the words used by the court provide governments with discursive resources to question the validity of a domestic policy, and push for the legitimacy of compliance. As above, one particularity of the trade regime is thus that the “audience” that ultimately requires persuasion is made up of domestic constituents. When addressing these constituents, 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 government thus form an unlikely alliance: together, they must persuade domestic audiences of the legitimacy of compliance.

As the introduction to this Special Section stresses, international courts have no way to compel their targets to follow their rulings, which leads judges to engage in legitimation politics. The oft-quoted phrase in the case of the trade regime points out that the WTO has neither truncheons nor tear gas.\textsuperscript{21} 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

\textsuperscript{20} Kono 2006.
\textsuperscript{21} Bello 1996.
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.

In our telling, the WTO employs greater affect alongside other strategies for dealing with political controversy already identified in existing studies. For instance, 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, only this strategy is not addressed to the same audience as the deployment of affect. And WTO panelists may still seek to circumscribe their rulings by relying on techniques like judicial economy, that allows them to limit the number of findings against the respondent government.22 But net of such issue-avoidance, the court will also seek to offer governments discursive resources to help them deal with domestic politics.

More generally, this article contributes to the view that adjusting the language of judicial reasoning, beyond the direction of the ruling, is one means by which judges handle political controversy. In a recent article, Dunoff and Pollack (2017) described the conflicting interests judges face by referring to a "judicial trilemma," whereby the design of courts can pursue at most two of the following three core values: judicial independence, judicial accountability (via control by political actors), and judicial transparency, by which the authors mean the identification of opinions with specific judges. The WTO's Appellate Body, in this telling, is a high-accountability court, with judges appointed for short, renewable terms. We agree with Dunoff and Pollack's ground premise that judges seek to hold states accountable to the commitments they have made, while remaining aware of the potential for political pushback; judges pursue "credibility, acceptability, and legitimacy, combined with the paramount concern for independence."23 We also share their assumption that judges are sufficiently politically savvy to recognize which rulings will be

---

perceived as most politically sensitive. What we put forth in this article is that there are means other than unsigned opinions that courts can use to walk the fine line between judicial independence and the political constraints they face. Among these, courts can fine-tune the type of language they rely on in formulating their opinions, to offer resources that pro-compliance constituencies in governments can use to argue for the validity of the ruling. In this way, the deployment of affect-laden language, not to mention established techniques like the exercise of judicial economy, are means of expanding the confines of the judicial trilemma.24

The necessary premise behind our argument is that judges retain considerable discretion over whether to deploy affect or not, regardless of the issue they are ruling over. They can thus choose to take strong positions using affect-laden terms, or they can remain behind the "mask of law,"25 appealing to past precedent. The extent of discretion judges retain in the WTO was aptly illustrated by the ruling in EC—Asbestos. The Asbestos dispute was brought against the European Union by Canada, which lost on all the main claims it put forth. The majority ruling of the AB turned on a purely competitive relationship between asbestos fibres and PCG fibres (a less dangerous alternative to asbestos), examining end uses to determine whether asbestos and PCG fibres were "like products" under Article III:4. A concurring opinion, a rare occurrence in WTO jurisprudence, agreed with the ruling's direction, but took issue with its reasoning. The wording in the concurring opinion was telling by its contrast:

"PCG fibres, it may be recalled, have not been shown by Canada to have the same lethal properties as chrysotile asbestos fibres. [...] It is difficult for me to imagine what evidence relating to economic competitive relationships as reflected in end-uses and consumers'
tastes and habits could outweigh and set at naught the *undisputed deadly nature* of chrysotile asbestos fibres.”

Perhaps more remarkable than the concurring opinion itself is the fact that in a 75-page AB ruling, the remainder of the AB managed never once to mention the terms "lethal" or "deadly", or any equivalent, when discussing asbestos. It took the concurring opinion to do so, twice in one paragraph. The substantive thrust of the concurring opinion illustrates the distinction that interests us here: the same ruling could be arrived at from behind the mask of law, keeping to a strictly legal discussion—well-vested in precedent—of the competitive test to assess likeness; or it could take a value-laden stance on the matter at hand: the lethal and deadly nature of asbestos, and France's right to ban it. The majority opinion took the first route, while the concurring opinion pointed to the second, in the process relying on the kind of value-laden language we have in mind here. Our point is that when justifying this loss at home, these two judicial approaches are likely to appeal to different audiences. When the court seeks to provide governments with the resources to convey the legitimacy of the ruling to their general public, we argue that it will be more likely to err towards the second approach.

The choice over wording thus has consequences for how governments can present the ruling to their domestic audience. Recall that according to trade theory, even those government actors favorable to compliance may find it politically difficult to push through reforms in accordance with the WTO's recommendations. The premise of our argument is that WTO judges are mindful of these domestic political pressures, and in cases where those pressures are likely to be high, they will choose the wording of their rulings accordingly, in a way that provides governments with the discursive resources to argue for the legitimacy of compliance before their domestic audiences.

---

Consider another 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 modifying 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."27 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. What is striking in reading the COOL ruling is the judges' repeated insistence on these terms denoting damage. Findings of violation at the WTO do not depend principally on the effects of measures at issue. That is, a measure can be found in violation under WTO law even if it has no distortionary economic effect to speak of, just as a highly damaging measure can be found to be consistent, if it is applied even-handedly. Such insistence on the value-laden aspect of a measure is what interests us, but comparing such insistence from one case to another is difficult to do in a systematic fashion. In the empirical section, we rely on automated text analysis to assist us in this task. For our purposes, what the COOL ruling illustrates is how governments can use rulings to

---

push for the legitimacy of reform.

The US, as all governments on the losing end of a dispute, claimed to be disappointed with the outcome of the COOL 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."  

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 persistently 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 impose a disproportionate burden", that it had "a detrimental impact", that it imposed a "record-keeping burden", which in turn had additional "detrimental impact". 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". It is worth emphasizing anew that WTO law relies on findings of discrimination; it is largely unconcerned with the magnitude of distortion, or its welfare effects.

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, were also then in turn picked up by media aimed at the agricultural community— 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 actors 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 service 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. In our argument, affect plays an analogous function, by providing governments with discursive resources to draw on in arguing with domestic audiences for compliance—or, as with the EC—Asbestos case, for the validity of an unsuccessful complaint.

It is further notable that legislators in the episode over COOL explicitly identified the ruling's language as coming from the court. In general, legislators need little help in finding the right words to make their case—the discursive resources that courts provide are valuable because they

---

express validation or condemnation, yet still bear the imprimatur of the court, and thus draw on its authority. Hannah Arendt famously distinguished semantic authority as distinct from both coercion and persuasion through argument.\textsuperscript{32} Our own reasoning is analogous: when it comes to politically fraught rulings, the force of the WTO's rulings comes from the unusual combination of an authoritative body expressing itself not in the legalese it is usually associated with, but in affect-laden terms.

More broadly, the US debate over COOL is thus a striking example of the judicialization of international relations in practice. US Republican legislators—far from being 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."\textsuperscript{33}

In sum, while the outcome that interests us is the wording of legal rulings, which puts us squarely in the "adjudication politics" phase spelled out in this Special Section, courts adjust the

\textsuperscript{32} Arendt 1977, 93-94.
wording of these rulings with an eye to what comes next. In this way, adjudication politics is informed by judges' anticipation of how likely compliance with their verdicts is, and by the "feedback politics" that may follow. The prospect of such feedback will loom especially large in cases that touch on politically sensitive matters traditionally associated with state sovereignty. In these more recently judicialized areas, such as health and safety standards and other regulatory areas, the pushback is likely to come from a broad domestic audience concerned about the government's ability to regulate in the public interest. This is in contrast to cases over established rules dealing with fiscal matters, which are more likely to bear on import-competing firms with a better understanding of the relevant law.

Our contention is that while WTO panels are more likely to stick to straight legal reasoning when dealing with such fiscal matters, they become more likely to deploy affect when ruling on behind-the-border regulatory issues, such as in the case of COOL. How generalizable is the COOL case—that is, how common is it for courts to rely on affect 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 seek to examine in the empirical analysis.

2.2 The Cost 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, managing members' expectations, 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 also render their rulings less generalizable beyond the dispute at hand, and leave the court more open to accusations that it is straying from its mandate. While affect-laden language provides resources to those government actors pushing for compliance, it thus also leaves them prone to criticism. A large literature argues how the unique authority of courts comes precisely from how they are bound to the law. When judges use their discretion to rely on strong sentiment in recognition of domestic political incentives of member states, they do so at the risk of being seen as overly politicized. This is the recurrent challenge that all international courts face: they must be at once politically savvy, and avoid being perceived as playing at politics, since their authority flows precisely from their commitment to stick to the law. Affect-laden passages are likely to become focal points in a ruling, and may change an entire ruling's perception. Walking the fine line between anticipating political reactions and sticking to the law is what keeps judges from relying on affect in their rulings to too great an extent.

The second consideration is that of generalizability. Through their rulings, judges seek to shape jurisprudence. But 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 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. Addressing non-legal audiences come to the detriment of addressing legal audiences, who become less likely to pick up on an affect-laden ruling to justify their reasoning in a subsequent case. To use our initial Zeroing example, the AB appeared to 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 the interpretation of the AD Agreement. In this isolated case, the AB's rhetorical flourish appeared, at least initially, to have paid off. The affect-laden passage was unlikely to travel beyond the case in question, but following this third dispute over zeroing, the US announced it would finally comply with the sum total of findings against it.

Yet the same Zeroing case also functions as an illustration of the other side of the trade-off faced by judges. In one of the most controversial moves in WTO history, in May 2016, the US blocked an AB member's reappointment. The consensus among observers was that the US decision was motivated by the AB's unfavorable rulings on zeroing. While the AB member in question, Sueng Wha Chang, had not sat on the zeroing case cited above, and thus could not have been the author of that affect-laden passage (AB opinions remains anonymous, and so it is not possible to know which AB member wrote this opinion), the event was nonetheless an example of the US striking back against what it perceived to be an overly strong position by the AB on a controversial issue. While many in the US believe that zeroing should be abandoned as a policy, given the legal consensus against it, in this case the anti-compliance contingent prevailed. The greater point is that as much as deploying affect can be effective in providing discursive resources to pro-compliance actors, it may also make judges prone to attacks from anti-compliance actors if it is seen as straying beyond their mandate. Taking a value-laden stance also means foregoing the benefits of ruling behind the "mask of law".

2.3 Parallel Judicial Tools

35 See DSU Art. 14 and DSU Art. 17, which specify that individual panel and AB opinions "shall be anonymous."
In this article, we focus on the role that affect plays 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. Prior to formulating an opinion, for instance, 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. 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. In nearly half of all WTO disputes, panels exercise what is called judicial economy. 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. Yet the objective of 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

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,

---

36 Garrett, Kelemen and Schulz 1998, 150.
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.

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”, “conflict”, “insufficient”, ”infring”, ”irrelevant”, ”arbitrary”, ”breach”, ”impair”, ”burden”. Positive words have roots such as “agree”, “respect”, “relevant”, “benefit”, “adopt”, “rights”, “principl”, ”reasonab”, ”adequat”, ”cooperat”, ”useful”, ”pertinent”, to cite the most frequently used across WTO rulings. We sum the two to obtain a relative frequency of total affect. We look at total affect, rather than positive vs. negative affect (a more common exercise in applications with media studies) because we are interested not in identifying some bias of judges, but in characterizing the type of language employed: to what extent do judges rely on value-laden language, whether positive or negative? The two, moreover, co-vary to a great degree, and the results hold when we look at positive or

---

39 Apel and Grimaldi 2014.
negative affect terms independently.\textsuperscript{40}

The strength of text analysis methods is that they can quickly summarize a given aspect of a long text. The necessary drawback, however, is that it can be difficult to offer an intuitive sense of what a score of e.g. 0.05 affect corresponds to, and how it compares to a text of 0.1. The utility of text analysis and the difficulty of offering an intuitive sense of their meaning are two sides of the same coin: it is because a human coder cannot accurately gauge the affect content of a 100-page ruling that these measures are attractive in measuring systematic differences in discursive choices. But in an effort to provide a better sense for the meaning of affect, we construct two swatches of text that are composites of various rulings, and that include only affect terms found in the WTO caseload. These text samples are meant to illustrate different means of arriving at an equivalent ruling.\textsuperscript{41}

Low Affect:

If a measure has an impact on imported products, this will not be dispositive of a finding under Article 2.1. However, the particular circumstances of the case, that is, "the design, architecture, revealing structure, operation, and application of the technical regulation at issue," lead us to conclude that the measure is \textit{inconsistent} with Article 2.1, since its impact on imported products does not stem exclusively from a regulatory distinction.

High Affect:

A measure that has a \textit{detrimental} impact on imported products may nonetheless remain \textit{consistent} with a country's obligations. However, the present measure is not applied in an \textit{even-handed} manner, since it does not stem exclusively from a

\textsuperscript{40} Dictionary approaches also face the challenge of distinguishing between the valence of terms and their use in a sentence, to know that e.g., "the present measure is \textbf{not} applied in an \textit{even-handed} manner" denotes negative sentiment. Pooling positive and negative terms together is the least error-prone means of circumventing this challenge.

\textsuperscript{41} For consistency, we sourced the various parts of our composite samples from rulings on TBT Article 2.1, which is often paired with GATT III:4 claims.
legitimate regulatory distinction. Its detrimental impact on imported products thus reflects a means of arbitrary and unjustifiable discrimination that is prohibited under WTO obligations.

In these illustrative examples, the first, low affect text features a sentiment score of 0.014, while the second would be measured at 0.15. To tie this back to our sample of WTO rulings, compare these scores against Figure 1, which illustrates the distribution of affect for panel reports and AB reports, where the bulk of the variation is found between 0.04 and 0.1 affect. Figure 1 also shows a well-behaved normal distribution for both types of rulings, suggesting that the analysis is unlikely to be driven by outliers.

In Figure 2, we once again plot the distribution of affect across WTO rulings, this time sorting it 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.

[ Figure 1. “Distribution of Affect Across Rulings” about here ]
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 2: “Average Affect Across Agreements” about here ]

3.1.1 Focusing in on GATT III Cases

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 variation within 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 predicted that, without national treatment, Members would simply replace lost tariffs with domestic fiscal or non-fiscal measures. National treatment was thus meant to be an “anti-circumvention device,” 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. Since the logic of GATT, writ large, was to cap and cut MFN tariffs, convert all nontariff barriers to MFN tariffs and cut these too, the burden fell on national treatment to ensure the value of these tariff bindings. GATT III:2, first sentence, covers “like goods,” for which differential taxation is, on its face, protectionist. GATT III:2, second sentence, covers products that are “directly competitive or substitutable,” for which differential taxation above a de minimis 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

---

42 Horn and Mavroidis 2013
43 Zhou 2012.
44 Horn and Mavroidis 2013
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 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 inevitably engage in at least some affective framing in all cases. But 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, and risks being perceived as veering away from judges' mandate merely to interpret the law. 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.45 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 control for the affect

content of complainant submissions, the assumption being that any hardwired sentiment inherent to a given provision would appear in complainants' submissions, especially. Controlling for these, we can be more certain that we are observing judges' own strategic behavior. We then go on to 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 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) claim 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. It should be said that as opposed to our expectations over affect, this normalization effect of mounting jurisprudence is more likely to operate through elites, such as legislators and policymakers, who have knowledge of past cases and the accumulation of rulings, than through their domestic
audiences.

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 associated drop in the amount of affect deployed over time. The $n$-th 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.”\textsuperscript{46} About 2300 words within the Lexicoder sentiment dictionary are flagged at least once in the WTO texts.

Importantly, we confine our analysis to the \textit{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 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 \textit{EC—Customs Classification of Frozen Boneless Chicken Cuts}; \textit{EC—Selected Customs Matters}; and \textit{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 \textit{US—Softwood Lumber}, \textit{Mexico—High-Fructose Corn Syrup (HFCS)}, and \textit{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.

\textsuperscript{46} Young and Soroka 2012.
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 appear 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 in our second analysis 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.

3.2.1 Findings: Affect in National Treatment Claims

We confine our main 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 \( GATT \text{ III:4} \) indicator variable as being 1 for all GATT III:4 disputes, and 0 otherwise. GATT III:2 cases are a little less frequent, making up 30% of the sample. We collapse panel and AB rulings together, and include an AB dummy variable to distinguish between them. The AB may be more vested in the politics of compliance, and if Figure 1 is any indication, it may be more prone to affect-laden language as a result.

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. Since we are considering a single provision, the sample is necessarily small, which is why we keep these models
as parsimonious as possible. Yet if anything, the size of the sample makes the significance of the findings all the more striking. Indeed, looking at the bare-bones univariate regression in the first column, 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 our estimation. The first is the indicator for AB rulings. The second is a control variable for any disputes that include either superpower as a defendant. The concern might be that judges will be especially careful in their choice of language when dealing with powerful complainants, in the same way that some evidence points to judges being more likely to use judicial economy when dealing with the US or the EU—though we ultimately remain agnostic as to the direction of the effect. We also account for time using cubic splines at four knots, since we have reason to believe that reliance on affect may become less necessary over time, at least within provisions.

In the third model, we also control for the affect scores of complainant submissions in the dispute at hand. We think of these as the best proxies for any inherent affect that might be associated with one dispute over another. If despite this control, GATT III:4 cases appear to lead to more affect in rulings, then we can be more confident that the findings speak to strategic behavior on the part of judges.

Table 1: Affect in WTO National Treatment Rulings

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATT_III4</td>
<td>0.729***</td>
<td>0.630**</td>
<td>0.526**</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(0.261)</td>
<td>(0.209)</td>
</tr>
<tr>
<td>US or EU Defendant</td>
<td>0.462</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.327)</td>
<td>(0.295)</td>
<td></td>
</tr>
<tr>
<td>AB Ruling</td>
<td>0.584**</td>
<td>0.612***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.228)</td>
<td>(0.215)</td>
<td></td>
</tr>
<tr>
<td>Complainant Submission Affect</td>
<td></td>
<td>11.426**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.256)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.632***</td>
<td>8.408</td>
<td>6.074</td>
</tr>
</tbody>
</table>
### Table

<table>
<thead>
<tr>
<th>Cubic Splines</th>
<th>(0.130)</th>
<th>(16.019)</th>
<th>(12.028)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>r2</td>
<td>0.15</td>
<td>0.37</td>
<td>0.46</td>
</tr>
</tbody>
</table>

DV is affect in WTO rulings. Sample is limited to GATT III:2 and GATT III:4 that have been ruled on. Robust standard errors clustered on common dispute: * p < 0.10, ** p < 0.05, *** p < 0.01.

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 13% more affect on average. Interestingly, the AB does appear to rely on affect significantly more than the panel, to the rate of about 10% more affect, all else equal. And although the affective content of complainant submissions is unsurprisingly highly related to that of the rulings themselves, what is most striking is that even controlling for this proxy, we see a strong and significant effect of GATT III:4 cases on the affect content of rulings.

#### 3.2.2 Findings: Affect and Legal Novelty

Next, 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 2, the answer appears to be yes. Here, the sample is made up of claims that have been ruled on by either the panel or the AB. 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 to bear a significant negative relation to the amount of affect used. The more panelists and the AB rule on a given legal issue, the less affect-laden their reasoning becomes. We control for the same considerations as in Table 1: we include an indicator for AB rulings, and the presence of a superpower defendant, and model the time trend using cubic splines, as before. The time trend, in particular, allows us to say that the effect we observe is not reducible to passing years, as e.g., a court becomes more sure of itself. What matters is the number of prior rulings on

Table 2: Legal Novelty and Affect in WTO Rulings

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Claims</td>
<td>-0.085***</td>
<td>-0.067***</td>
<td>-0.044***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>AB Ruling</td>
<td>0.316**</td>
<td>0.309**</td>
<td>0.293**</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.127)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>US or EC Defendant</td>
<td>0.182</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.195)</td>
<td>(0.178)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.777***</td>
<td>17.617</td>
<td>20.336**</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(12.296)</td>
<td>(9.455)</td>
</tr>
<tr>
<td>Cubic Splines</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>749</td>
<td>749</td>
<td>749</td>
</tr>
<tr>
<td>r2</td>
<td>0.09</td>
<td>0.13</td>
<td>0.29</td>
</tr>
</tbody>
</table>

DV is affect in WTO rulings. Sample consists of all claims having received rulings across the WTO caseload. Column (3) includes fixed effects on legal claim. Robust standard errors clustered on common dispute:

*p<0.10, **p<0.05, ***p<0.01
a given issue. Because we are looking at multiple claims, we cluster all robust standard errors on the dispute in all estimation. And in model 3 of Table 2, we consider within-provision variation, by including claim-type fixed effects. This amounts to asking, do findings on e.g. antidumping lead to less affect being deployed in subsequent rulings over the same antidumping issues? The answer appears to be yes. The effect of legal novelty holds when we include respondent fixed effects, or control for the direction of the ruling at the claim level, or the number of total claims brought in the dispute. In all cases, more rulings over more established areas of law appears to be related with less on affect.

In sum, the findings offer us a coherent picture of judicial behavior in the face of political controversy. In disputes bearing on national sovereignty, judges deploy greater affect in an attempt to offer litigant governments discursive resources to convince their audiences of the legitimacy of their findings. But looking across the entire caseload, 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 during the phase of litigation politics, as the introduction to this Special Section outlines. They must hew to the word of the law while remaining mindful of the political sensibilities of sovereign states. In formulating their rulings, courts must keep an eye on the stages to come: compliance politics and feedback politics. The resulting tension is further 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 vary the clarity of their opinions.\textsuperscript{47} They can also defer to prior practice, by increasing citations to case law or external sources of authority.\textsuperscript{48} We argue that alongside these established judicial strategies, judges in the WTO, aware of the centrality of domestic politics, handle political controversy by deploying greater amounts of words denoting overt positive or negative sentiment.

We argue that this strategic judicial behavior 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 (or national) courts in charge of implementing rulings, legal opinions can address themselves principally to these other judicial actors. This unlikely alliance in the trade regime also speaks to the extent to which politics has been judicialized: when US Republican legislators invoke the words of the WTO to justify domestic reforms, we can infer—recent claims to the contrary notwithstanding\textsuperscript{49}—the extent to which international courts matter in domestic political debate. That these legislators focus on the court’s more affect-laden pronouncements also speaks to what legislators believe will hold the greatest sway with domestic audiences. The authority of international courts and affect-laden rhetoric: the combination speaks to the particular form of argument that results from judicialized politics. Political debate is in part transmuted into legal argument, yet the constraints of the political process remain: domestic audiences must be persuaded of the validity of compliance, and to this purpose, the mere authority of international courts is insufficient.

\textsuperscript{47} See Staton and Romero, this Special Section.
\textsuperscript{48} Hume 2006.
\textsuperscript{49} See Ginsberg and Abebe, this Special Section, for a fuller discussion of the various claims about judicialization and de-judicialization in international affairs.
It is likely that the particular alliance struck between courts and defendant governments varies from one country to another. We have assumed that governments need to persuade domestic audiences of the legitimacy of international courts' rulings, but we readily allow that this need will vary according to context. In this way, Staton and Romero (in this Special Section) argue that IACHR judges defer to defendants by modulating the vagueness of their rulings when they lack perfect insight into a country's implementation politics. Similarly, it might be that the deployment of affect also responds to country-specific characteristics. Most plainly, non-democracies might face different incentives vis-à-vis their domestic audiences. Future work could explore the extent to which the affect in WTO rulings responds to regime type, or the strength of the political opposition in a given case.

Whatever the national context, however, too much reliance on affect-laden terms by courts also limits the generalizability of findings, and leaves judges open to charges of having strayed beyond their mandate. Judges must seek to walk that line. This is why in the WTO, we see affect being deployed in cases where governments have the greatest need for discursive resources to argue for the legitimacy of compliance before their domestic audiences.


economic research.


Sauter, Oliver. 2014. Communicated Uncertainty of the ECB. In *Monetary Policy under Uncer-
tainty. Springer pp. 177–197.


