Are WTO Rulings Biased?
The Role of Institutional Design in Protecting Judicial Autonomy

Joost Pauwelyn[†] Krzysztof Pelc[‡]

Abstract

The dispute settlement system of the World Trade Organization prides itself on its high degree of judicial independence and the impartiality of its adjudicators. Yet compared to other international tribunals, WTO members exert considerable political control over WTO adjudicators. Contestation over appointments of adjudicators also reflects governments' awareness that nationality may in fact matter for outcomes. *Does it*? An empirical analysis of 25 years of Appellate Body activity offers a nuanced answer. Exploiting the random allocation of adjudicators to AB divisions, we find no evidence of systematic national bias looking across the board. Yet we do find evidence of bias for AB chairs, suggesting that when adjudicators are singled out, they become more prone to political pressure. A similar effect pertains to individual dissenting opinions: the presence of a co-national on a division is associated with significantly increased odds of dissent. Judicial independence at the WTO has long been taken for granted. Our findings suggest that such trust is largely warranted, yet that even small tweaks in institutional design increase political pressure on adjudicators, in ways that threaten impartiality. This holds significant implications for WTO reforms going forward.

[†] Graduate Institute of International and Development Studies, Geneva and Georgetown University Law Center, Washington, DC. joost.pauwelyn@graduateinstitute.ch
[‡] Corresponding author. Department of Political Science, McGill University. kj.pelc@mcgill.ca.
We thank Sean Nossek and Marie Fester for excellent research assistance.

1. Introduction

The authority of courts rests on their perceived independence. In the case of international law, impartiality is essential in securing buy-in from member states, and from domestic audiences which most directly bear the cost of complying with adverse rulings. The World Trade Organization is no exception in this respect. Until the US began blocking all further appointments to the WTO's Appellate Body in 2019, this tribunal had been looked to as an exemplar of the successful judicialization of international politics.[1] This success rests in large measure on the widespread perception that the AB was impartial in its verdicts. WTO adjudicators themselves often loudly proclaimed this independence. As James Bacchus, one of the institution's first Appellate Body members, and a US-appointee, noted early on: "there has *never once* been a suggestion by any Member of the WTO that the Appellate Body is anything but independent and impartial."[2]

This is largely true. Be that as it may, both formal WTO rules and state behavior reflect concerns about how the nationality of adjudicators may influence outcomes. WTO panelists from the litigant countries, or even from those countries that join as third parties to the dispute, are thus barred from serving on a case, unless by explicit approval of the parties—which is almost never granted. And while the same rule does not apply at the Appellate Body level, the reason for this also reflects concerns about national "representation" on the bench: given that the US and the EU participate in a disproportionate number of WTO disputes (both as complainant and defendant), a rule disallowing AB members from ruling on cases involving the country that appointed them would have deprived the trade system's two traditional superpowers of representation in a majority of disputes, something that neither member accepted. Similarly, from the very start of the WTO, the US and the EU demanded that a seat on the AB be permanently reserved for one of their co-nationals, and although this demand was never inscribed in the treaty texts, it was always honored.[3] In sum, there is at once a consensus that national bias has no place in the institution, and a recognition that having national representation among adjudicators is highly valuable.

[1] See, among others, Goldstein et al (2000), and the ensuing literature on the legalization of world politics, for which the passage from the GATT to the WTO served as the primary case of study.
[2] Bacchus 2003. Original emphasis.
[3] Elsig 2013. In fact, as Elsig and Pollack (2014) demonstrate, the level of Member scrutiny over the (re)appointment process of AB members significantly increased over time.

The result is that AB members (as opposed to panelists) regularly find themselves ruling on cases involving their country of origin. This leads us to ask two simple questions: does the record of rulings reflect any national bias? And what might the presence or absence of such bias depend on? Now is an especially apt moment to be posing these questions. Appointments to the AB are currently blocked by the US, following claims that the AB overstepped its mandate, and the appeal stage of the WTO judicial function is thus paralyzed. Negotiations over WTO reforms of the dispute settlement understanding have entered a pivotal phase, with a hoped for outcome in time for the 2024 Ministerial Meeting.[4] Yet if it comes back into service, it is likely that the AB will be of a different form. Specifically, likely reforms will offer member states more oversight. This leads us to consider the 25-year record of the AB, and draw out any lessons that might apply to ongoing reforms.

In so doing, we contribute much-needed empirical evidence to a longstanding debate. Despite the general perception of WTO dispute settlement as unbiased, developing countries have occasionally denounced the AB as "pro-American," starting with the initial selection of members at the WTO's inception.[5] More recently, the administration of Donald Trump repeatedly made the opposite claim, decrying the "anti-U.S. bias" of the dispute settlement body.[6] These competing claims have usually been assessed by counting wins and losses and comparing them across states.[7] Such exercises may be descriptively useful, yet they are also methodologically flawed, and may thus end up doing more harm than good.

Bias is notoriously difficult to detect. A growing literature shows how bias can be entirely implicit, or unconscious, rendering the accounts of adjudicators themselves of limited use.[8] From a methodological standpoint, moreover, tests of bias have to contend with the issue of selection into litigation: the cases involving a given country, like the US, are likely to be fundamentally different from those involving other countries, and any differences in outcomes may be due to these *ex ante* differences, rather than any *ex post* skew in decision-making by adjudicators. Two

[4] Politico. May 2023. "Reform or Die? If the US gets its way, the WTO might do both". https://www.politico.eu/article/reform-die-usa-washington-world-trade-organization-wto-ngozi-okonjo-iweala-joe-biden/

[5] "The WTO, its secretariat and bias against the South". Third World Network Info Service on WTO and Trade Issues. https://www.twn.my/title2/wto.info/2019/ti190420.htm

[6] "U.S. Keeps Winning WTO Cases, Despite Claim of Anti-U.S. Bias" https://news.bloomberglaw.com/international-trade/u-s-keeps-winning-wto-cases-despite-claim-of-anti-u-s-bias

[7] Ibid.

[8] See Danziger et al 2011. The literature on bias is most developed when it comes to racial bias. In the context of judicial decisions, see Rachlinski et al. 2008.

aspects of the institutional design of the WTO allow us to overcome this methodological challenge and improve on simple counts of countries' wins and losses: (i) AB members can rule on matters involving their home country, and (ii) they are appointed to disputes by a process of random selection.

Our theoretical expectation, which builds on recent studies on institutional design, is that insofar as bias is present, it would be a reflection of political pressure—either tacit or explicit—on the part of member states. This expectation leads to a testable implication: bias should be more observable in settings where individual adjudicators are singled out, that is, where they lack the political cover offered by the three-member division. When this occurs, adjudicators should find it harder to resist pressure to give in to states' demands.

The first way in which adjudicators are singled out is through the appointment of a chair of the ruling division, or a "presiding member." Presiding members have more influence over the proceedings; they meet with the parties to discuss working procedures, and are often the panelist with the greatest experience. Past work shows that the identity of panel chairs matters more for outcomes than the identity of the other two panelists (Busch and Pelc 2011). Such a perception should increase political pressure on those individual adjudicators. A second source of variation occurs across time: initially, AB members imposed a strong norm of consensus opinions on themselves, which meant the division of three would always rule together. This norm was successfully maintained during the first decade of the WTO, but it then began to erode. As it did, political pressure increased: now that adjudicators were seen as able to dissent, they faced greater expectations that they would. This relates to the setting in which individual adjudicators are most singled out, namely in dissenting opinions, which are by definition single-authored. Although these are nominally anonymous, precisely in an attempt to shield adjudicators from political pressure, recent studies have shown that this anonymity is largely make-believe. As a result, insofar as bias is present, we argue that it is especially likely to manifest itself in dissenting opinions.

In testing these expectations, our analysis presents a nuanced set of findings. On the basis of 25 years of rulings on appeals, and exploiting the random assignment of AB members to different cases, we find no evidence of national bias in the AB's opinions, on average. We run a power analysis to ensure that this null effect is not due to sample size or other aspects of the estimation. Looking specifically to US adjudicators, given how the US has been at the center of the clash between political control and judicial independence at the WTO, we once again fail to find consistent evidence of bias by US AB members in favor of the US. What this finding also implies

is that AB divisions that lack a US judge are not as a result more likely to rule against the US. That is, the system, absent US representation, does not appear any more *un*favorable to the US.

These findings are a testament to the self-professed independence of the WTO's highest tribunal. This lack of national bias, by itself, distinguishes the WTO from other international judicial settings, like the International Court of Justice (ICJ) and a number of international arbitral settings, where consistent findings have produced strong evidence that judges favor the states that appoint them (Posner and Figueiredo 2005; Puig 2019). Alongside nationality, we also examine whether traits such as adjudicators' professional backgrounds, their gender, and whether they come from a developing country origin matters for outcomes. Interestingly, of all these personal traits, only gender appears to have a consistent effect: the presence of women is significantly associated with more findings of violation.

While the overall absence of national bias in the AB is thus notable, we do see signs of variation across time and setting. First, when we look at disputes that occurred post-2005, we find that having a co-national from the defendant country among the adjudicators does appear to decrease findings of violation. As we suggest, this inflection corresponds to the point when a theretofore strong norm of consensus rulings among AB members began to erode. Adjudicators became more exposed to political pressure once they were perceived as free to dissent from the majority ruling.

We then look at the nationality of the chair of the sitting three-member division. Here, we find evidence of bias across the caseload. When the chair of the ruling AB division happens to be from the defendant country, the proportion of findings favorable to the complainant drops by 29%. This effect, which is even more pronounced when we look specifically at the US, is consistent with our broader argument. Whenever adjudicators are singled out—as when they are appointed chair of a division—they lose the protection offered by the collective aspect of the three-member division. They are more exposed to political pressure. Autonomy comes under threat; bias results.

Finally, we ask whether dissents become more likely when one of the AB members is from the respondent country in a given case. The blunt answer is yes: the odds of a dissent increase more than fivefold when one of the AB members is from the defendant country, even once we account for dispute characteristics. If we focus only on so-called "true dissents,"[9] the effect is starker still:

[9] According to the definition set out in Kim and Mavroidis (2018).

the odds of a dissenting opinion go up tenfold when one of the division's adjudicators is from the respondent country.

Taken together, these findings present a coherent picture of the AB on the basis of the full universe of its legal rulings. The absence of bias on average across the caseload, in contrast to the evidence of national leanings by AB division chairs and authors of dissenting opinions, speaks to how a unified group of even three adjudicators can remain sufficiently immune to pressures from WTO Members. By contrast, any setting where adjudicators are singled out increases political pressure. In this way, the data speak to the importance of institutional design in safeguarding judicial independence.

2. Theoretical Background

An enforcement body that would be overly deferential to powerful political actors, bending to their preferences, would fail at its primary function of reducing uncertainty; governments would be the first to lose out.[10] Yet the same governments are wary of delegating too much power to courts, lest adjudicators show insufficient regard for political exigency, or worse, run away with their mandate and start modifying the meaning of the concessions states have made to one another. If they lack a means of curbing such perceived excesses, governments may refuse to delegate power to an enforcement body in the first place.

A symmetrically opposite dilemma exists from the standpoint of adjudicators: their judicial authority rests on being perceived as impartial and immune to political pressures; yet they must remain sufficiently attuned to the sensitivities of Member-states, lest that authority be openly disregarded or forcibly taken away by disgruntled governments.[11] Most recently, and in the specific tribunal we look at here, Kucik and Puig (2022) claim that Appellate Body members in the WTO adapt their prior precedents when confronted with noncompliance from governments. In this way, adjudicators from different international tribunals share a common set of incentives: they seek to offer legal opinions that reflect their interpretation of the treaty as applied to the facts of the case; they want to see these rulings meet with compliance from governments; and most pragmatically, they want to maintain their appointments, and protect their subsequent career prospects.

[10] Guzman (2005).
[11] Garrett, G., Kelemen, R.D. and Schulz, H., 1998. Staton and Moore (2011). Ferejohn (2002). Alter (1998).

As a result of these competing incentives, the greater the level of political control over a tribunal, the more likely that, when faced with politically sensitive issues or precedent-setting questions, adjudicators will defer to powerful political actors, since they will have internalized the possible consequences of ruling against the national interests of member-states.[12] The result is the familiar tension between judicial independence and political control that characterizes all international judicial settings. Domestic judicial settings must manage the same tension in their institutional design (Staton and Moore 2011). Yet given how international public law rests inexorably on the consent of sovereign nations, the exertion of political control in the international realm (especially in the name of national interest) usually draws less condemnation from domestic audiences. The WTO is a case in point. As a result, the challenge that adjudicators face in balancing judicial independence with the political sensibilities appears especially acute in the international realm. In this article, we are interested in how successful the WTO, and especially its once-celebrated Appellate Body, has been at walking this fine line.

A hortatory commitment to judicial independence is present throughout the WTO treaties,[13] and is often reiterated by the adjudicators themselves. We open this article with a quote to this effect from an AB founding member. Here is another from the last sitting AB member, Hong Zhao of China, in her farewell speech to the institution, in 2020: "I am constantly reminded of my duty as an Appellate Body Member to settle disputes for WTO Members independently and impartially."[14]

Yet despite these guarantees and repeated assurances, governments themselves have demonstrated a constant concern over national representation on the AB. The selection of the initial AB bench of seven adjudicators, out of a list of 32 candidates, was hotly contested, and gave rise to allegations of manipulation by the US from developing countries. India was the loudest critic,

[12] Garrett, Kelemen and Schulz (2002).

[13] DSU Article 17:3, sentence 2 provides that the AB's members "shall not be affiliated with any government." At the establishment of the AB, the Dispute Settlement Body expanded on this requirement, linking it explicitly with the notion of judicial independence, under the heading of "Impartiality". *Establishment of the Appellate Body. Recommendations by the Preparatory Committee for the WTO approved by the Dispute Settlement Body on 10 February 1995*.

[14] Hong Zhao Farewell Speech. 2020. https://www.wto.org/english/tratop_e/dispu_e/farwellspeechhzhao_e.htm. Similarly, as Lacarte-Muró declared at the end of his final term as the first Chair of the AB: "We are well aware that none of our rulings is likely to be greeted with universal approval; but our function is another: to be independent, impartial and objective at all times. I believe this also to have been the case." (Farewell Address to the Dispute Settlement Body. 19 December 2001).

but Brazil also "expressed systemic concerns regarding the composition of the Appellate Body."[15] Even Switzerland had misgivings about national representation.[16] The US and EU had initially claimed two seats each on the bench, and then curtailed their demands in the face of strong opposition from the rest of the membership.[17] Yet their demand to have a *de facto* permanent seat on the AB was acceded to by the membership for the next 25 years.[18] Even this was not seen as sufficient for the EU, considering the combined weight of EU countries in the global trading regime. At an earlier meeting, the EU representative had vowed to "submit a proposal for the [subsequent] Singapore meeting to ensure a more representative membership for the Appellate Body and one that would adequately reflect the role of the EU in the multilateral system."[19]

Governments thus plainly imbue the nationality of adjudicators with considerable importance, all the while publicly recognizing the utmost impartiality of those adjudicators. This apparent paradox reflects the conflicting incentives of governments within any third-party enforcement body: governments rely on the system's perceived legitimacy, yet they seek to maximize their advantage within it.

The WTO is far from exceptional in seeking some measure of representativeness by putting criteria on the nationality of ruling adjudicators. The benches of the European Court of Justice and the European Court of Human Rights feature one representative from each of their member states. Slots of the International Court of Justice bench are distributed regionally. Given the size of its membership, the WTO cannot aspire to full representativeness, and its adjudicative body of seven is thus necessarily skewed towards some countries over others. Does this skewness matter for outcomes?

---

[15] In a veiled reference to the US and EU, India observed: "The Committee appeared to have been influenced by considerations not related to these four [agreed upon] elements. It appeared that while most delegations had been asked which candidates they considered the most appropriate appointees to the Appellate Body and why, some delegations seemed to have the privilege of answering a question as to which candidates they objected to. However, they seemed to have no obligation to answer why. India regretted that this 'extraordinary privilege' enjoyed by a few delegations seemed to influence the final outcome decisively." WTO DSB Minutes of Meeting Held in the Centre William Rappard on 1 and 29 November 1995. WT/DSB/M/9.

[16] As the Swiss representative noted, "the composition proposed at the present meeting did not correspond with some basic criteria stipulated in the DSU. A definition of the European entity was too restrictive: i.e., it represented an erroneous approach of economic and political realities and therefore was not correctly reflecting a '*representative balance*.'" Ibid, emphasis added.

[17] See, e.g. "WTO Establishes Appellate Body." Nov 30, 1995: https://www.sunsonline.org/trade/process/followup/1995/11300095.htm. The same article also observed, recognizing the clashing incentives of governments and adjudicators, that the way in which "the United States was effectively given the 'privilege' of objecting to some names, and thus helping to label some of the successful ones as "pro-American"" represented "a burden that none of the seven [chosen adjudicators] would appreciate."

[18] Elsig 2013.

[19] Supra at note 14.

2.1 The Elusive Nature of Bias

A common approach to the question of judicial bias is to count wins and losses. These exercises were frequently offered in an attempt to rebuff claims from the US during the Trump administration about an anti-US bias in the AB. As one such news analysis ran, "The Trump administration argues that the WTO's Dispute Settlement System is biased against the United States. Yet […] one count finds that the U.S. wins more than the average when it is complainant, and loses less than the average when it is respondent." That report referred to another which found that "the U.S. has a better-than-average rate of success in arguing cases at the WTO."[20]

Yet such counting exercises, while well-intended, are inherently flawed. The main reason is a problem of selection into litigation. When descriptive statistics show the US winning "more than the average when it is a complainant,"[21] (a difference which, it is worth noting, was not statistically significant at conventional levels) that average rests on a strong assumption, namely that the type of cases that the US brings is comparable to others. It might still be that foreign adjudicators are biased against the US, or American adjudicators are biased for it, but that the US brings fundamentally different types of cases. In fact, we know this for a fact: in an early assessment, Bown (2013) found that while developing cases bring more legally straightforward cases over tariffs and trade remedies, developed members like the US bring more complex cases over matters of regulation, or non-violation claims, which have less certain legal outcomes. If this is the case, then even a rate of wins and losses that is exactly equal to the average could hide significant favorable bias. Other factors that are correlated with nationality may also sway outcomes. We know that cases brought by developing countries, especially early on, attracted more third parties—other member states that join the deliberations and offer their own submissions—which are known to lower the odds of settlement (Busch and Reinhardt 2006). That, in turn, results in a pool of cases with higher merit. Conversely, if the US is able to use extra-legal means to push for concessions without having to resort to dispute settlement, then it will be left with a pool of lesser legal merit, having dismissed the "easier" cases before ever filing for consultations. Some of these factors are observable, and can thus be accounted for in a regression setting; yet countless others are unobservable factors which are more difficult to account for. Most plainly, it may be

[20] Supra at note 5.
[21] Ibid.

that the weight of countries' submissions is a reflection of their legal capacity: some countries may simply argue their position better than others, having more resources to devote to the case. The result is that simple counts risk hiding the truth rather than revealing it.

Luckily for analysts, two aspects of the design of the WTO offer a means around this challenge of selection into litigation: there is no provision against AB members ruling on disputes involving their own countries; and the process of allocating AB members to disputes is random, approximating the setting for a natural experiment.[22] Indeed, as the WTO itself outlines, "The process for the selection of Divisions is designed to ensure randomness, unpredictability and opportunity for all Members to serve regardless of their national origin."[23] The sequence of events is important: litigants first decide whether to appeal the panel decision, and then they find out which adjudicators from the AB are assigned to the case. As Article 6(2) of the Working Procedures sets out, "The Members constituting a division shall be selected on the basis of rotation, while taking into account the principles of random selection, unpredictability and opportunity for all Members to serve regardless of their national origin." It bears mention that this random appointment process is a design feature we owe to the AB itself, rather than WTO Members, who did not specify how AB members would be appointed to particular cases, having given the AB *carte blanche* to set its working procedures. The exact lottery system used is kept secret, but its purpose and intended outcome are known. The former AB member A.V. Ganesan has referred to the "ingenious random selection mechanism devised by the Appellate Body."[24] The result would then amount to a quasi-natural experiment, where each dispute is either randomly "treated" with the co-national adjudicator treatment, or gets the non-national "control." It is then possible to ask whether this co-national "treatment" has an observable effect on outcomes, without fearing that the estimation is contaminated by, for instance, countries' influence over appointments to particular cases, or their decision to appeal given knowledge of who will be deciding the case.

By contrast, the same exercise could not be performed on WTO panelists, who are picked by mutual agreement of the litigants themselves from a shortlist of *ad hoc* adjudicators proposed

[22] To our knowledge, Arias (2018) is the first to point out this analytical benefit in a working paper. One challenge in assessing political pressure in any tribunal comes from how states have every reason to act strategically. Since states have some insight into adjudicators' priors, they have an incentive to sway the process of appointment of adjudicators to particular cases, as happens in the appointment process of WTO panelists. In such circumstances, it becomes impossible to distinguish potential judicial bias at the individual level from bias resulting from influence applied to some appointments over others.

[23] "Appellate Body Members". https://www.wto.org/english/tratop_e/dispu_e/ab_members_descrp_e.htm

[24] Ganesan, *supra*, fn 12.

by the Secretariat. This process has grown increasingly fraught; so much so that in 64% of all disputes,[25] countries have been unable to agree on a set of panelists. In those cases, the WTO Director General, alongside Secretariat staff, have had to assign one or more panelists. That decision, in turn, likely reflects various strategic interests of the institution, and thus cannot be considered random. Perhaps for this reason, a strict rule explicitly prohibits co-nationals of either litigant party (as well as third parties) from serving on panels (unless both parties agree), in an effort to limit the influence of powerful litigants over the appointment process. This formal provision eliminates a source of bias, but it means that the most frequent litigants have few co-nationals shaping jurisprudence, which is the main reason why the Appellate Body does not have an analogous constraint. The result is that analysts have no means of reliably verifying whether panelists are swayed by the litigants' nationality. Yet most of the public debate over bias has concerned the AB, rather than the first-stage panel decisions. We are thus able to assess the possibility of bias in the most setting where the question has been most relevant.

The random assignment of AB members can itself be verified. This is a necessary check to carry out, since individual AB members conceivably have an incentive to seek appointment on some cases over others. Indeed, some anecdotal evidence suggests that there are ways for adjudicators to get around the random selection process: knowing the likely order of cases in the pipeline, they can declare a conflict of interest, or make themselves unavailable over a given time period, in a way that could skew the assignment. To preview this finding, which we detail below, we find no evidence of any bias in the selection of AB adjudicators to decide specific cases. On every parameter we examine, the selection appears as-if random, as per the AB's own working procedures. This allows us to go on to ask whether a given AB division is more likely to rule in favor of a litigant country if it is "treated" with an adjudicator from that country.

2.2 Variation in the Exposure to Political Pressure

We hold theoretical priors about the political pressures AB members must grapple with. The premise underlying our expectations is that judges are more exposed to pressure from their home country than from other Member-States. And whenever individual adjudicators are singled out, we argue, they become more vulnerable to political control, and their autonomy suffers as a result.

[25] Pauwelyn and Zhang. 2018.

The most common form of political control is pressure by Members on their co-nationals. The US allegedly taking steps against the reappointment of its own AB members is a case in point. To be sure, pressure on co-nationals is not the only form that political pressure can take. After all, the WTO's consensus rule allowed the US to eventually block the reappointment of Seung Wha Chang, a Korean AB member. Yet insofar as governments find it easier to exert pressure on the co-national they appointed—because of shared professional networks, or employment prospects following their tenure at the WTO—this offers us an opportunity to compare different settings for signs of such pressures. Settings with heightened political control should be where we observe more evidence of national bias.

It is this type of political pressure, we argue—rather than inherent nationalist sympathies, or the internalization of the values of their home state—that is likely to lead to bias. As a result, absent institutional protection, AB members may be more inclined to rule in favor of their home state. If political pressure is at play, this pro-national inclination should be more obvious in settings where adjudicators are more isolated. Ours is thus a story of institutions, rather than one of individual-level psychology.[26] We believe that adjudicators' commitment to impartiality is genuine; the issue is that as recent studies highlight, the design of the DSU, owing chiefly to short terms and the reappointment process, leaves WTO adjudicators more exposed to political pressure than in any comparable international tribunal.[27]

We focus on three aspects of the DSU where political pressure may be most apparent. This first occurs in the case of the "presiding member" elected by each division, who is commonly called the division chair. The role of chairs has been shown to matter at the panel level. Busch and Pelc (2010) find that whereas the experience of the average panelist does not affect legal outcomes like the rate of appeal, the experience of the panel chair *does* matter.[28] In our case, whether division chairs actually have greater influence over the content of the ruling, it is enough for governments to believe that they do.[29] By being singled out, those individual adjudicators would then be more likely to be held responsible by governments, and would have reason to act accordingly. When

[26] For the latter, see e.g. Puig 2019, who lists a set of cognitive biases that adjudicators may be prone to.
[27] See Dunoff and Pollack 2018; Günther 2020.
[28] Busch and Pelc. 2009.
[29] This may be a reasonable assumption looking at the arbitration setting, which is the other judicial context that relies on a chair in a three-person panel. There, chairs sometimes serve an umpire-like function, and can decide a case if the three-member panel is unable to converge on a decision.

political interests find a point of pressure, this should observably affect outcomes. This leads us to separately test for national bias of AB division chairs.

The second aspect we examine is variation through time. Early on, in a notable move, under the rubric of "consensus" and "collegiality", adjudicators imposed a strict norm of on themselves, which provided them some additional political cover: they vowed not to render separate opinions, even though the treaty texts explicitly allowed them to, and they vowed to consult with the other four members on the bench in all their decisions. As the first chair of the AB put it, "from the very beginning, I felt strongly that we should avoid minority opinions at all costs."[30] This belief was institutionalized in the AB's own working procedures, which read: "The Appellate Body and its divisions shall make every effort to take their decisions by consensus. Where, nevertheless, a decision cannot be arrived at by consensus, the matter at issue shall be decided by a majority vote."

For the first decade of the WTO, AB members steadfastly held to this norm; not a single dissent was issued, and this exceptional record of consensus was often highlighted by AB members themselves.[31] Then, starting in 2005, complaints about the strictures of the norms of consensus and collegiality started to emerge. Later, one adjudicator bemoaned "an over-emphasis on 'collegiality' that shaded into peer pressure to conform" (Graham 2020). As the norm of consensus eroded, and eventually fell apart, separate opinions grew more common. As they did, member governments could suddenly expect adjudicators, and especially their co-nationals, to issue dissenting opinions. Specifically, countries in the defendant seat could exert pressure on a co-national sitting on the AB to dissent from the majority opinion, which in a super-majority of cases finds against the defendant. This variation across time also proves analytically valuable, since it allows us to test whether national bias becomes any more common after 2005, once the political cover offered by the norm of consensus was weakened.

The third aspect of dispute settlement we consider is the related matter of dissenting opinions themselves. Once the norm of consensus disappeared in its strong form, pressure to issue favorable dissents would have grown. Although dissenting opinions are unsigned, the identity of a given dissent's author is a matter of active speculation within Geneva circles.[32] Recent work using text

[30] Julio Lacarte-Muró. *Launching the Appellate Body*.
[31] Ehlermann, an AB member, noted how the Working Procedures had "clearly prevailed over the possibility" of dissent offered by the formal rules Ehlermann (2003). In 2009, Alvarez wrote that "the functioning of the Appellate Body (AB) is virtually perfect in terms of collegial decision-making," and the means of assessing the perfection of the record was by gauging the prevalence of unanimous decisions (Alvarez 2009).
[32] Ehlermann (2002). Some WTO observers have gone so far as to conclude that "everyone involved" knows who the author of an AB dissent is (Terris et al, 2007.)

analysis shows that it is possible to pinpoint the most likely author of a dissent using authorship detection tools; yet most Geneva insiders may not need such sophisticated means. This lack of *de facto* anonymity leads us to ask whether co-nationals were more likely to issue dissents favorable to their home country, when that country was one of litigants.

As Dunoff and Pollack (2018) argue, in the face of strong political control, judicial autonomy suffers when legal opinions can be attributed to individual judges. We look for empirical support for this idea, looking not only at dissents—which are Dunoff and Pollack's focus—but also more innocuous design features like the appointment of a "presiding member" to each ruling division. We compare these aspects of institutional design to some commonly examined personal traits of adjudicators. Chiefly, there is reason to believe that the professional background of AB members could matter for outcomes: for instance, it is often said that panelists are pulled from the ranks of trade officials because these are more likely to be understanding of governments' political sensibilities. By contrast, lawyers may be more wedded to an independent interpretation of the law. We also test whether gender matters in any way, since a growing number of studies has alluded to the under-representation of women on the bench of international tribunals, in ways that may matter for outcomes.[33]

3. Empirical Analysis

What can the twenty-five-year record of the AB tell us about the competing forces that AB members are exposed to? Specifically, how has the judicial autonomy that adjudicators have consistently laid claim to fared under varying levels of political pressure?

Our empirical analysis has three main parts. We begin by assessing the purported randomness of the appointment process; we then look for signs of co-national bias among all AB members, and AB division chairs in particular; and we end by looking at how the presence of co-national adjudicators affects the odds of dissenting opinions. As per our discussion above, we expect greater bias whenever adjudicators are singled out—either as divisions chairs, or as the writers of dissents.

Our various tests rely on a common dataset, which builds on Horn and Mavroidis (2020) and Kucik and Pelc (2018). We update the data to cover the entire existing AB caseload, and add

[33] See the Symposium introduced by Maučec and Dothan (2022) on the effects of international judges' personal characteristics.

information about dissenting opinions and the tenures of individual AB members. Our main dependent variable of interest is the direction of the ruling, which we code at the claim-level, and aggregate as a dispute-level ratio, corresponding to the proportion of claims ruled in favor of the complainant, given all the claims on which a decision was reached. Our main unit of analysis is thus the dispute, since all our explanatory variables, such as the identities of the adjudicators, are also observed at the dispute level.[34]

3.1 Assessing Random Assignment

As noted above, the AB's Working Procedures explicitly set out that the appointment of individual adjudicators to cases be random. Yet anecdotal evidence suggests that there are means of manipulating the process. Because the DSB rules include provisions against conflicts of interest, adjudicators may cite such a conflict, or simply claim not to be available for a particular appeal, as a way of increasing their odds of being appointed to the next case down the line. States can also time their own appeals strategically, knowing which AB members have already been assigned to prior appeals, so as to increase their odds of having a particular adjudicator appointed to their case. Given these considerations, does the empirical record offer any evidence of such manipulation?

To test this question, we create an original dataset of all AB members selected for any given case. Every dispute counts seven possibilities, except for a handful of exceptions in the recent past, when the number of AB members began to dwindle, and the number of possible picks grew smaller. All told, our dataset counts 835 dispute-adjudicators. For each dispute, we observe which of the available AB members were selected to a given division. In a first preliminary test, we compare the balance between the "treated" and control groups, that is, those divisions with a co-national adjudicator and those without. We look at several covariates: the number of claims in a case, the number of resulting rulings, the proportion of claims in favor of the complainant, and the length of the panel proceedings. None show any sign of imbalance between the groups. We then look more closely at nationality, where the expectation of manipulation would be strongest. We estimate how a correspondence between an adjudicator's nationality and that of either litigant affects the odds of that adjudicator being selected to the case. In other words, if a case concerns

[34] In our robustness checks, we rerun the analysis at the claim level, and our findings remain entirely unchanged. In our baseline estimation, the effect on a ruling of violation of having a AB member from the defendant (complainant) party on the ruling division is negative (positive), yet neither approaches conventional levels of significance.

country X, is an adjudicator from country X any more (or less) likely of being appointed to the dispute?

Table 1 presents the results. None of our estimations yield any statistically significant effects: we find no sign of manipulation. Neither the Co-national of Respondent variable nor the Co-national of Complainant variable shows any effect. We then focus on the US and the EU, given that they are the DSB's central players, its most active litigants, and have, by tacit agreement, co-national representatives sitting on the AB at all times. We code a US and EU indicator variable for all US or EU AB members, and interact this variable with a dummy indicating a US or EU defendant or complainant. We then check whether any individual adjudicators are more likely to be chosen for specific legal issues raised in the dispute at hand. Finally, we check whether any national origin seems more likely to be selected, on average, paying special attention to the US and EU. We observe no effect here, either (see Appendix).

We follow up these estimations by running a similar test of the randomness of AB chair appointments. Here, we verify the odds of an AB member from a given country to be elected as "presiding member." There is no pre-specified formal means of selecting the division chair, though there is anecdotal evidence that at least in some cases, the adjudicators on the division have made the choice by flipping a literal coin. When we examine the resulting odds of assignment, the US AB members do not seem significantly more likely to be elected presiding members, conditional on being appointed to a given division.[35] The higher number of US co-nationals serving as division chairs ruling on cases where the US is a respondent (as happened on 14 occasions, against 2 for the EU) is thus a reflection of the number of cases where the US is a respondent, rather than the odds of being elected as presiding member. In sum, on all relevant observables, AB members seem to be assigned to divisions in an as-if random fashion, and the same appears true of the appointment of AB chairs—though it bears repeating that in the latter case, we cannot be as confident in the nature of the actual selection process. Overall, these results speak to the procedural legitimacy of the AB's assignment method, which also allows for a cleaner test of our next question of interest.

[35] More specifically, conditional on the American AB member being among the three adjudicators on a division, the proportion of cases in which that member is elected is 37.8%, which is not significantly different from chance. In a more pointed test, we verify whether the US being among the litigants makes it more (or less) likely that the US AB member will be elected chair of a division. And here, too, the results are statistically insignificant: whether the US is the complainant, the defendant, or one of the two, the odds of the US member becoming the division chair are not different, statistically speaking.

Table 1: Is the Assignment of ABMs to Divisions Random?

	(1)	(2)	(3)	(4)
ABM Co-national of Respondent	-0.066			
	(0.162)			
ABM Co-national of Complainant	-0.183			
	(0.160)			
US Co-national		-0.007	-0.007	0.024
		(0.167)	(0.157)	(0.136)
US Respondent		-0.013	-0.013	
		(0.052)	(0.102)	
US AB member X US Respondent		-0.069	-0.069	
		(0.275)	(0.252)	
EU Co-national		0.065	0.065	0.134
		(0.148)	(0.137)	(0.146)
EU Respondent		-0.082	-0.082	
		(0.065)	(0.140)	
EU AB member X EU Respondent		0.278	0.278	
		(0.388)	(0.349)	
US Complainant				-0.030
				(0.126)
US AB member X US Complainant				-0.341
				(0.328)
EU Complainant				-0.026
				(0.112)
EU AB member X EU Complainant				-0.104
				(0.284)
Constant	-0.105**	-0.121**	-0.121+	-0.126+
	(0.026)	(0.042)	(0.071)	(0.065)
Observations	835	835	835	835

Estimates from probit regression. Dependent variable is an indicator of whether an eligible AB member is selected to a given AB division. Robust standard errors clustered on common dispute in parentheses + p<0.10 * p<0.05 ** p<0.01.

3.2 Assessing Rulings

Having confirmed that the assignment of adjudicators to cases appears random, we can now ask whether the presence of a co-national AB member has an observable effect on legal outcomes. As per our discussion, we are interested in this test as a window into whether, and when, political pressure is felt most strongly by adjudicators. Our expectation is that whenever adjudicators are singled out, they become more vulnerable to such pressures.

Table 2: Co-national AB Members from Litigant Countries

		AB Member from Complainant Country		
		0	1	Total
AB Member from Defendant Country	0	94	34	128
	1	44	8	52
	Total	138	42	180

How often do AB members sit on cases that concern their own countries of origin? As noted above, this possibility is expressly disallowed at the panel level (unless both parties agree, which rarely happens). But as Table 2 shows, it is a frequent occurrence at the AB. There were 44 disputes that counted AB members from the defendant country, 34 disputes that counted AB members from the complainant country, and 8 more cases that featured an ABM from both the complainant and the defendant country. All told, at least one co-national from a litigant country sits among the three ruling ABMs in 48% of cases. While this seems a high number, it is in fact dictated by the random assignment to divisions, combined with the tacit agreement by which the system's two most frequent users always have a co-national on the AB bench. It also makes plain how many adjudicators would need to be excluded if the AB had chosen, as it could have done, to bar co-nationals from serving on disputes involving their countries of origin.

Table 3: Rulings and AB Co-nationals

	AB member from Respondent Country	Number of disputes	Proportion of findings in favor of complainant	Standard Error
All Respondents	0	127	0.720	0.030
	1	50	0.624	0.055
US Respondent	0	41	0.695	0.051
	1	31	0.646	0.068
EU Respondent	0	15	0.716	0.066
	1	13	0.436	0.109

Next, we have a look at the simplest descriptive statistics linking the co-national "treatment" and AB ruling outcomes. Table 3 shows the average direction of the ruling, coded as the proportion of findings favorable to the complainant, according to whether a co-national was among the adjudicators.[36] It then focuses specifically on cases where the US and EU, who most frequently have co-nationals among the ABMs, are defendants. The results are mixed. In all cases, the co-national treatment is linked to a more favorable ruling for the defendant country. The relationship is not statistically significant when looking at all defendants, or when focusing on the US. In the EU case, however, the association is suggestive of some pro-national leanings: the average ruling against the EU is 72 percent in favor of the complainant when no European AB member is present; but that falls to 44 percent when an EU AB member is present. This difference, moreover, is statistically significant in a simple t-test at the 0.05 level. Yet given how we are dealing with a relatively small number of rulings, this association could be due to confounding variables. The issues at stake in these disputes differ, there may be changing trends across time, or some cases may have been more politically sensitive than others, in ways that sway the results. Similarly, since these are appeals, they are in dialogue with a panel ruling at the prior stage. Next, we try and account for these confounding factors.

To do so, we rely on a series of generalized linear models (GLM) with a logistic link function, which are well suited to estimating a dependent variable that takes the form of a proportion bounded from 0 to 1.[37] We begin with a parsimonious estimation that looks at the effect of an ABM member from either the defendant or complainant country on the direction of the

[36] Where the denominator is made up of all findings delivered by the AB in a given case. More than one finding may be delivered for a given claim brought by a litigant.
[37] Papke and Wooldridge. 1996.

ruling, controlling only for time trends, using cubic splines with three knots. To try and account for the prior stage in the dispute settlement process, we control for the proportion of claims ruled in favor of the complainant by the panel. Most cases are ruled in favor of the complainant at the panel stage; although the AB frequently amends or reverses panel findings, rulings remain highly favorable to complainants net of appeal. Controlling for that first ruling, we want to know whether an appeal is more likely to be ruled in favor of the complainant if the AB features a co-national.

In subsequent models, we add control variables for the most contentious disputed legal issues, namely antidumping, subsidies and countervailing duties, SPS, TBT, and agriculture disputes. To try and capture the systemic implications of the dispute, we control for the number of third parties. Third parties are other Members that ask to join the proceedings, and who can make oral and written submissions which are included in the final report.[38] We cluster robust standard errors on the common dispute, to account for how some rulings relate to more than one complaint.

The findings in Table 4 are consistent across our various specifications. While the effect of having a co-national among the AB adjudicators ruling on a case has a consistent positive effect for that country, it never reaches statistical significance. That is, the presence of an AB member from the respondent country is associated with fewer findings of violation; conversely, the presence of an AB member from the complainant country is associated with more findings of violation. But neither rises to conventional levels of statistical significance. In columns 5 and 6, we then look specifically at the US and EU, by interacting our indicator for AB member from the respondent country with a US/EU dummy. Here again, the presence of EU adjudicators shows more of an indication of pro-national bias than US adjudicators, but the effect falls just short of statistical significance. The controls behave as expected. Third parties, in particular, are associated with more pro-complainant findings, in support of existing work.[39] In Column 7, we add a set of four variables covering personal traits of AB members: trade policy background, legal training, gender, and developing country origin. These do not appear to have an effect.

In the last column (8), we run the same estimation on the subset of disputes that arose after the norm of consensus began to erode, with the appearance of dissenting opinions in 2005. Here, we begin to see signs of bias: a co-national adjudicator from the defendant country appears to be related to fewer findings of violation (by comparison, no such effect is visible in the WTO's first

[38] The presence of third parties has been argued to be one way by which adjudicators grasp the political implications of their recommendations. See: Busch and Reinhardt. 2006. Brutger and Morse. 2015.
[39] Johns and Pelc 2018.

decade, prior to 2005). This is consistent with our argument about the AB's self-imposed norms providing political cover. Yet looking at the overall history of the institution, and especially its early years, these results nonetheless offer support for Bacchus's emphatic contention: "We are always independent. We are always impartial. We will always be."[40]

One possible concern is the size of the sample, which is necessarily limited by the number of AB rulings. Recent work has suggested that much of quantitative political science may be underpowered (Arel-Bundock et al. 2023). This is also a concern for null findings. Although Arel-Bundock (2023) imply that there is an under-reporting of null findings across the literature,[41] their findings also suggest that it is difficult to distinguish null findings from would-be positive findings that are plagued by lack of power. In our case, the question is whether the lack of evidence of bias is an indication of judicial autonomy at the WTO, or simply an artifact of a (necessarily) small sample. To help address this concern, we perform a power analysis on our baseline estimation.

Table 4: The Effect of AB Member Co-Nationals on AB Rulings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Respondent Co-national on AB	-0.515		-0.529		-0.239	-0.701	-0.506	-1.053*
	(0.375)		(0.385)		(0.459)	(0.637)	(0.385)	(0.615)
Panel Ruling Direction	1.854**	1.799**	1.910**	1.953**	1.951**	1.973**	1.991**	2.243**
	(0.503)	(0.512)	(0.558)	(0.559)	(0.578)	(0.567)	(0.599)	(0.810)
Complainant Co-national on AB		0.210		0.225				
		(0.324)		(0.327)				
EU Respondent					-0.059			
					(0.605)			
Resp Co-national on AB X EU Resp					-0.933			
					(1.035)			
US Respondent						0.246		
						(0.495)		
Resp Co-national on AB X US Resp						0.171		
						(0.926)		
Sum Trade Official ABMs							-0.037	
							(0.339)	
Sum Lawyer ABMs							0.254	
							(0.381)	
Sum Women ABMs							0.478	

[40] Supra, fn 2.

[41] Given a pervasive problem of low power, most analyses should report null findings. The fact that they do not speaks to potential selection for positive findings.

							(0.352)	
Sum Developing Country ABMs							0.330	
							(0.289)	
Agriculture			-0.348	-0.195	-0.294	-0.267	-0.418	0.351
			(0.472)	(0.459)	(0.525)	(0.497)	(0.463)	(0.689)
Antidumping			0.251	0.395	0.303	0.198	0.256	-0.268
			(0.439)	(0.424)	(0.455)	(0.482)	(0.422)	(0.580)
SCM			0.065	0.083	0.056	0.012	-0.000	0.468
			(0.354)	(0.353)	(0.365)	(0.381)	(0.351)	(0.597)
SPS			-0.058	-0.103	0.052	-0.014	0.090	-0.459
			(0.492)	(0.466)	(0.551)	(0.544)	(0.509)	(0.757)
TBT			-0.149	-0.114	-0.039	-0.204	-0.183	-0.094
			(0.405)	(0.424)	(0.427)	(0.415)	(0.427)	(0.807)
Number of Third Parties			0.033	0.028	0.024	0.029	0.036	-0.030
			(0.032)	(0.030)	(0.034)	(0.032)	(0.029)	(0.068)
Cubic Splines	X	X	X	X	X	X	X	X
Observations	179	179	179	179	179	179	179	69

GLM regression estimates. Dependent variable is the proportion of findings ruled in favor of the complainant. Robust standard errors clustered on common dispute in parentheses * p<0.10 ** p<0.05 *** p<0.01. Sample in model (8) is restricted to post-2005 AB rulings.

This requires us to determine an expected effect size. There is no objective means of doing so, yet theory can provide some guidance. The average number of claims in our sample is 10.6. Our main dependent variable is a proportion of findings in favor of the complainant. We determine that a relevant bias effect would correspond to |0.1| in a linear regression, or one additional finding, either in favor of the defendant or the complainant, as a result of national bias. We are also guided by a literature on judicial behavior that systematically finds that judges are swayed by political pressures from member states (Carrubba et al. 2008). The power calculation suggests that we have sufficient power to identify an effect of 0.10 if it were present in the data, assuming a (standard) power level of 0.80. Specifically, a sample size of 162 would suffice; the sample size of our baseline test is 179. That said, the sample remains small, and thus a necessary caveat to our ability to affirm definitively that AB adjudicators are unbiased. Yet using all the available evidence, we find no sign of consistent bias in either direction, looking at the average adjudicator.

One can perform further spot checks to see the relation between co-national AB members and ruling outcomes in specific salient cases. For instance, one might think that a country like China would be especially likely to exert pressure on the adjudicators it appointed. There are only two cases in which the Chinese AB member sat on the ruling division in a challenge against China

(China—*Rare Earths* and China—*GOES*), which is too few to draw any systematic inference. Yet it remains telling that in both these cases, the AB ruling upheld all claims against China brought by the complainants: even adjudicators appointed by a country like China, which is often criticized for making the rule of law subservient to state interests as interpreted by the Central Committee of the Chinese Communist Party, were able to exercise sufficient judicial autonomy to rule against their country of origin in two key disputes, without issuing a dissent. The suggestion is that judicial independence can stand up to political interests in the international realm.

Yet as the last model in Table 4 shows, this ability may rely on successfully hewing to judicial norms of behavior—when these begin to erode, as they did in 2005, signs of possible bias appear. As we demonstrate next, judicial independence also depends on other aspects of institutional design.

Table 5: The Effect of AB Co-National Chairs on AB Rulings

	(1)	(2)	(3)	(4)	(5)
Div Chair from Defendant Country	-0.932**	-0.973*	-1.201**	-1.041*	-2.955**
	(0.465)	(0.525)	(0.536)	(0.566)	(1.212)
Panel Ruling Direction	1.956***	2.054***	3.132***	2.189***	2.119**
	(0.508)	(0.550)	(0.816)	(0.555)	(0.858)
Agriculture		-0.299	-0.779	-0.343	0.541
		(0.492)	(0.567)	(0.491)	(0.672)
Antidumping		0.281	0.363	0.332	-0.071
		(0.428)	(0.499)	(0.432)	(0.605)
SCM		0.022	-0.663	0.058	0.545
		(0.364)	(0.450)	(0.379)	(0.643)
SPS		-0.190	0.156	-0.245	-0.178
		(0.477)	(0.564)	(0.532)	(0.720)
TBT		-0.199	0.141	-0.118	-0.308
		(0.434)	(0.673)	(0.513)	(0.870)
Number of Third Parties		0.031	0.006	0.033	-0.074
		(0.031)	(0.037)	(0.031)	(0.070)
Div Chair Trade Official				-0.070	
				(0.441)	
Div Chair Lawyer				-0.318	
				(0.528)	
Div Chair Woman				1.344**	
				(0.617)	

Div Chair from Developing Country				-0.182	
				(0.452)	
Observations	179	179	179	179	69
AB Member Fixed Effects			X		
Time Cubic Splines	X	X	X	X	X

GLM regression estimates. Dependent variable is the proportion of findings ruled in favor of the complainant. Model (5) sample based on post-2005 disputes. Robust standard errors clustered on common dispute in parentheses
* p<0.10 * p<0.05 ** p<0.01.

To recall our central expectation, whenever adjudicators are singled out, they lose some of the protection afforded by a collegial body, and thus become more prone to political pressure. One of the ways in which this happens is through the selection of one of the three AB members as division chair. Are such chairs more likely to be swayed by national interests? To test this, we rely on a similar set of estimations as above. This time, our explanatory variable is an indicator of whether the AB division chair is from the defendant country. The results are presented in Table 5, and they tell quite a different story from Table 4. The presence of an AB chair from the defendant country is strongly associated with fewer findings of violation against the defendant. The size of the effect is similar across the estimations in columns (1)-(3): the average ruling is 29% more favorable to the defendant when the division chair is from the defendant's country, holding all else equal. In column (4), we add our battery of AB member attributes, this time looking only at ruling division chairs. Of these, one is notable: when women are division chair, findings of violation appear significantly more likely. Finally, we once again restrict the sample to the post-2005 period, when we argue the AB lost some of the political cover provided by a commitment to avoiding dissenting opinions. Here too, the effect of having a division chair from the respondent country is compounded, and findings of violation appear significantly less likely.

Additionally, we take a closer look at the US, the one country with enough division chairs to allow us to estimate an effect.[42] As before, we interact the Co-national Division Chair with a US Respondent variable. Tellingly, the effect is even more pronounced than for other countries: a US chair is associated with 33% fewer findings against the US, all else equal. When we then look at cases where the US is a complainant, we find the opposite: a US chair is associated with 41% *more* findings of violation, and the effect is strongly significant.

[42] See the discussion above, in Section 3.1, of the observable randomness of AB chair appointments.

The difference between the findings in Tables 4 and 5 speaks to the premise of our argument. One might think that whether a co-national is present on the AB division or is the "presiding member" should not make much difference. Yet by singling out an individual adjudicator, the role of presiding member offers an opening for political pressure to bear down on the AB. The effect is strong enough to be observable across the caseload. The prime instance in which individual adjudicators are singled out, however, is through the option of issuing dissents.

3.2 Assessing Dissents

In examining dissenting opinions, we rely on the Kim and Mavroidis (2018) coding, as well as their characterization of "true dissents" as opposed to all separate opinions. The outcome of interest is now binary, so we run a probit model. Throughout, we control for the direction of the AB's majority opinion, since this is what the dissent is most clearly reacting against. Cubic splines are included to account for time, an especially important aspect of the story, given the erosion of the consensus norm through the AB's history. In Table 6, we begin by considering the odds of any separate opinion in columns 1 and 2; we focus on true dissents in columns 3 and 4.

Throughout the various estimations of Table 6, dissenting opinions of all sorts are strongly and significantly related to the presence of a co-national from the defendant country among the adjudicators. The magnitude of these effects is large. The odds of a separate opinion of any kind become 6 times greater in the presence of a co-national AB member. The odds of a "true dissent" grow 14 times higher when a co-national is present on the division, all else equal. This is consistent with how dissenting opinions overwhelmingly favor defendants. We interpret these findings as an indication of how political pressure is especially high on co-nationals in terms of dissenting opinions. While the three-person ruling division offers co-national adjudicators political cover, the option of individual dissents does not: governments come to expect it in cases of negative majority opinions. To paraphrase the above-cited judge from the ECJ, because dissents allow adjudicators to show their home country what a loyal judge they are, they feel pressure to do so.[43]

[43] Terris et al. 2007, supra.

Table 6: Estimating the Odds of AB Dissenting Opinions

	(1)	(2)	(3)	(4)
	Separate Opinions		True Dissents	
Respondent Co-national on AB	0.895***	1.207**	1.373**	2.629**
	(0.345)	(0.483)	(0.569)	(1.123)
Majority AB opinion direction	0.415	0.908*	1.002	2.331***
	(0.567)	(0.516)	(0.650)	(0.901)
Panel Ruling Direction		-0.674		-3.794***
		(0.983)		(1.198)
Agriculture		-0.230		-1.153
		(0.642)		(1.240)
Antidumping		0.462		-0.262
		(0.594)		(0.652)
SCM		1.122**		2.666***
		(0.480)		(0.958)
SPS		1.179*		
		(0.649)		
TBT		0.634		
		(0.568)		
Number of Third Parties		0.041		0.173*
		(0.032)		(0.098)
Observations	179	179	179	179
Time Cubic Splines	X	X	X	X

Probit estimates. Dependent variable is the probability of separate opinion (columns 1-2) or dissenting opinion (columns 3-4), according to the criteria set out in Kim and Mavroidis (2018). SPS and TBT indicators not included in column (4) because they do not vary on outcome. Robust standard errors clustered on common dispute in parentheses + p<0.10 * p<0.05 ** p<0.01.

4. Conclusion

The forced interruption of the AB's activities has brought to a halt the most active branch of the global trade regime. By holding trade enforcement in abeyance, it has injected uncertainty into international trade cooperation. Yet this inflection point also provides a natural opportunity for observers to look over the AB's quarter-century history and see what its record can tell us about the functioning of international tribunals. In particular, the decision by a superpower to block the enforcement arm of an organization that it was instrumental in creating underscores the

fundamental tension between judicial autonomy and political control that all tribunals must contend with. It is this tension, and its empirical manifestations, which we explore in this article.

While the judicial independence of AB members is often celebrated, often by the AB members themselves, members have occasionally leveled accusations of bias against the AB. The Trump administration denounced the AB as anti-US; earlier, developing countries had made the opposite claim, criticizing the AB appointment process as leading to pro-US bias. This wrangling over AB appointments, in itself, speaks to how member states must believe that having their nationals represented on the AB bench is somehow beneficial. These competing claims have most often been addressed by simply counting up a given country's wins and losses. As we argue, such descriptive exercises are misleading, given factors affecting the selection of cases into litigation.

We address this methodological challenge by exploiting the random allocation of AB members to divisions. When we do so, we find no evidence of bias on average, looking at the caseload as a whole. Yet this changes when we focus on settings where adjudicators were singled out in one way or another. Specifically, while the average adjudicator does not appear swayed by nationality, this does not seem true of division chairs, who we argue are more exposed to political pressure. The average ruling is thus 29% more favorable to the defendant when the division chair is from the defendant's country, all else equal. Looking specifically at the case of the US, the presence of an American division chair is associated with 39% fewer findings against the US. Conversely, when the US is a complainant, a US chair is associated with 41% *more* findings of violation.

We also find evidence of variation across time. During the first decade of the WTO, the AB held to a strict norm of consensus rulings, which offered those adjudicators some political cover, by rendering responsibility more diffuse. When this norm began to erode, in 2005, this strength in numbers was lost, and pressure from governments increased. This turning point is associated with an observable difference in outcomes: we note growing signs of national bias post-2005, compared to the first decade of the WTO. Relatedly, the setting where adjudicators are most singled out is that of dissenting opinions. Here, the effect of nationality proves decisive: the odds of dissents grow up to 14 times higher when a co-national is present on the division. This set of three empirical tests offers a coherent picture of potential sources of bias in judicial decision-making, and a better sense of the conflicting pressures the AB must grapple with.

The findings hold considerable implications for WTO reforms going forward. It is highly likely, given US demands, that if the AB is ever to come back into service, member states will

have to exert greater political control over the proceedings. The risk is that such reforms may go too far. The dispute settlement system's independence has thus far been taken for granted—our findings suggest such trust is largely warranted. Yet we also demonstrate how quickly bias can appear, once a tribunal's institutional design is tweaked even slightly. This risk should be kept high in mind when seeking to correct the AB's perceived shortcomings, lest reforms come at the expense of the trade enforcement system's most valuable asset: its impartiality.

References

Alter, K.J., 1998. Who Are the "Masters of the Treaty"? European Governments and the European Court of Justice. *International Organization*, pp.121-147.

Alberto Alvarez-Jimenez. 2009. The WTO Appellate Body's Decision-Making Process: A Perfect Model for International Adjudication? *Journal of International Economic Law* 12(2), 289–331.

Anderson, C.J., 2007. The end of economic voting? Contingency dilemmas and the limits of democratic accountability. *Annu. Rev. Polit. Sci.*, *10*, pp.271-296.

Arias, Eric. 2018. "Impartiality in International Courts: Evidence from A Natural Experiment at the WTO". Presentation at the 2018 International Political Economy Society.

Bacchus, 'Table Talk: Around the Table of the Appellate Body of the World Trade Organization'. 2002. 35 Vanderbilt Journal of Transnational Law 1021, 1038.

Bacchus, James. 2003. *"Woulda, Coulda, Shoulda": The Consolations of WTO Dispute Settlement.* Address to the International Bar Association Geneva, Switzerland March 20, 2003.

Barfield, C.E., 2001. Free Trade, Sovereignty, Democracy: Future of the World Trade Organization. *Chi. J. Int'l L.*, *2*, p.403.

Brutger, Ryan and Julia C. Morse. 2015. Balancing law and politics: Judicial incentives in WTO dispute settlement. *The Review of International Organizations* 10(2): 179-205.

Busch, Marc L. and Eric Reinhardt. 2006. "Three's a crowd: Third parties and WTO dispute settlement." *World Politics* 58(3): 446-477.

Busch, Marc L. and Krzysztof Pelc. 2009. "Does the WTO need a Permanent Body of Panelists?". *Journal of International Economic Law* 12(3): 579-594.

Carrubba, C., Gabel, M., & Hankla, C. (2008). Judicial Behavior under Political Constraints: Evidence from the European Court of Justice. *American Political Science Review, 102*(4), 435-452.

Creamer, Cosette D. and Zuzanna Godzimirska, 2016. 'Deliberative Engagement within the World Trade Organization: A Functional Substitute for Authoritative Interpretations', 48 *New York University Journal of International Law and Politics* 413.

Danziger, S., Levav, J. and Avnaim-Pesso, L., 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), pp.6889-6892.

Duch, R.M. and Stevenson, R.T., 2008. *The economic vote: How political and economic institutions condition election results*. Cambridge University Press

Dunoff, J.L. and Pollack, M.A., 2018. The judicial trilemma. *The American Journal of International Law*, *111*(2), pp.225-276.

Ehlermann, Claus-Dieter. 2002. Six Years on the Bench of the "World Trade Court". Some Personal Experiences as Member of the Appellate Body of the World Trade Organization. *Journal of World Trade* 36(4): 605–639.

Ehlermann, C.D., 2003. Reflections on the Appellate Body of the World Trade Organization (WTO). In *Proceedings of the ASIL Annual Meeting* (Vol. 97, pp. 77-86). Cambridge University Press.

Elsig, M. and Pollack, M.A., 2014. Agents, trustees, and international courts: The politics of judicial appointment at the World Trade Organization. *European Journal of International Relations*, *20*(2), pp.391-415.

Elsig, Manfred. 2013. "The EU as an Effective Trade Power? Strategic Choice of Judicial Candidates in the Context of the World Trade Organization." *International Relations*, 27(3), pp. 325-340.

Ferejohn. John 2002. "Judicializing Politics, Politicizing Law. *Law and Contemporary Problems*" 65(3): 41-68.

Garrett, G., Kelemen, R.D. and Schulz, H., 1998. The European Court of Justice, national governments, and legal integration in the European Union. *International organization*, *52*(1), pp.149-176

Goldstein, J., M. Kahler, R.O. Keohane, and A.M. Slaughter (2000) 'Introduction: Legalization and world politics', *International Organization* 54(3), 385–399.

Graham, Thomas. 2020. Farewell speech of Appellate Body member Thomas R. Graham. WTO. https://www.wto.org/english/tratop_e/dispu_e/farwellspeechtgaham_e.htm

Günther, P., 2020. "Groupthink Bias in International Adjudication". *Journal of International Dispute Settlement* 11(1): 91-126. See the discussion of US AB member Tom Graham's speech, infra note **Error! Bookmark not defined.**.

Guzman, A.T., 2005. The Design of International Agreements. *European Journal of International Law*, *16*(4), pp.579-612.

Guzman, A.T., 2008. International tribunals: a rational choice analysis. *u. Pa. l. Rev.*, *157*,

Jackson, John. 2000. “Dispute Settlement and the WTO: Emerging Problems”. In: McWhinney, E. and Secretariat, W.T.O., 2000. *From GATT to the WTO: The Multilateral Trading System*. Kluwer Law International BV. 71.

Janow, Merit. 2008. *Reflections on Serving on the Appellate Body*. 6 Loy. U. Chi. Int'l L. Rev. 249.

Julio Lacarte-Muró. 2015. *Launching the Appellate Body*. Policy Commons. WTO.

Kim, Evan Y. and Mavroidis, Petros C., Dissenting Opinions in the WTO Appellate Body: Drivers of Their Issuance & Implications for the Institutional Jurisprudence. 2018. Robert Schuman Centre for Advanced Studies Research Paper No. RSCAS 2018/51. http://dx.doi.org/10.2139/ssrn.3293777

Kucik, Jeffrey and Sergio Puig. 2022. “Do International Dispute Bodies Overreach? Reassessing World Trade Organization Dispute Ruling.” *International Studies Quarterly.* 66(4).

Lacarte-Muró, Julio. 2015. “Launching the Appellate Body.*”* In *A History of Law And Lawyers in the GATT/WTO: The Development of the Rule of Law in the Multilateral Trading System*, ed. Gabrielle Marceau. Cambridge, 476-481.

Lacarte-Muró. 2005. ‘WTO Appellate Body Roundtable’, in Laurence R. Helfer and Rae Lindsay (eds), *New World Order or A New World Disorder?* Washington: American Society of International Law. p.183.

Lamp, Nicolas. 2021. *Arrested Norm Development: The Failure of Legislative-Judicial Dialogue in the WTO*. Working Paper, Queen’s University.

Koremenos, B., 2007. If only half of international agreements have dispute resolution provisions, which half needs explaining?. *The Journal of Legal Studies*, *36*(1), pp.189-212

Koremenos, B., Lipson, C. and Snidal, D., 2001. The rational design of international institutions. *International Organization*, *55*(4), pp.761-799.

Matsushita, M., 2005. Some thoughts on the Appellate Body. In *The World Trade Organization: Legal, Economic and Political Analysis* (pp. 1389-1403). Springer, Boston, MA.

Papke, L. E. and J. Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11: 619–632.

Pauwelyn, Joost and Weiwei Zhang. 2018. Busier Than Ever? A Data-Driven Assessment and Forecast of WTO Caseload, *JIEL*. p.18.

Pauwelyn, Joost and Krzysztof Pelc 2022. WTO Rulings and the Veil of Anonymity. *European Journal of International Law*. Forthcoming.

Posner, E.A. and De Figueiredo, M.F., 2005. Is the international court of justice biased?. *The Journal of Legal Studies*, *34*(2), pp.599-630.

Powell Jr, G.B. and Whitten, G.D., 1993. A cross-national analysis of economic voting: taking account of the political context. *American Journal of Political Science*, pp.391-414.

Puig, S., 2019. Debiasing International Economic Law. *European Journal of International Law*, *30*(4), pp.1339-1357.

Rachlinski, J.J., Johnson, S.L., Wistrich, A.J. and Guthrie, C., 2008. Does unconscious racial bias affect trial judges. *Notre Dame L. Rev.*, *84*, p.1195.

Simmons, Beth and Allison Danner. 2010. "Credible Commitments and the International Criminal Court". *International Organization,* 64(2): 225-256

Staton, Jeffrey and Will Moore. 2011. "Judicial Power in Domestic and International Politics." *International Organization*.

Terris et al, 2007, 124. Terris, D., Romano, C.P. and Swigart, L., 2008. Toward a Community of International Judges. *Loy. LA Int'l & Comp. L. Rev.*, *30*, p.419.)

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and Biases. *Science*, 185(4157), pp.1124-1131

Voeten, E., 2008. The politics of international judicial appointments. *Chi. J. Int'l L.*, *9*, p.387.

Voeten, Erik. 2008. "The Impartiality of International Judges: Evidence from the European Court of Human Rights." *American Political Science Review* 102(4): 417–33.

Wolff, A.W., 2001. Problems with WTO dispute settlement. *Chi. J. Int'l L.*, *2*, p.417.