Selecting for Ambiguity in International Dispute Settlement

s

Tonya L. Putnam
Department of Political Science
Columbia University
tp2202@columbia.edu

Prepared for the IPES Annual Conference
Duke University
November 11-12 2016

ABSTRACT

Deterring violations of international commitments is a key reason why states delegate power to international courts and other dispute settlement institutions. However, the more credible the enforcement threat posed by a tribunal the less it will be used to sanction blatant violations of legal rules. To the contrary, high capacity tribunals are more often preoccupied with resolving ambiguities in the applicability, scope, or content of legal rules. They do so by deciding disputes – but not just any disputes: highly concrete disputes over whether applicable law exists, or what it requires among parties with direct stakes in the outcome. They are, in short, *elaborating the law*. Rule elaboration and enforcement each involve distinct logics for drawing parties into legal disputes, and for selecting disputes into adjudication. Thus, lumping them together can lead to faulty inferences about the intentions and capabilities of disputing parties, and inflate estimates of non-cooperative behavior. It likewise obscures a critical function of judicialized tribunals at the international level. This paper explores the conditions under which recourse to third-party dispute resolution is likely to reflect the logics embedded in each concept using WTO dispute settlement cases on subsidies and countervailing measures. The results underscore that separating out the rule elaboration and enforcement functions of international dispute settlement bodies lends itself to a more richly strategic view of state behavior within and around these bodies.

1. Introduction

The concept of enforcement – the ability to compel observance of rules and obligations -- is critically important to the politics of international relations and international law. Absent the expectation that governments generally will honor their states' legal commitments and to adhere to customary rules, cooperation in the international system would be rare. Where blatant decisions not to follow international legal rules are costly to complying states, those failures not only undermine the agreement in question, but may also damage interstate relations and deplete resources for other cooperative efforts. And yet, the possibility that commitments will not be upheld is a consistent concern of international politics. Where individual states have acknowledged incentives to cheat – often or occasionally – achieving and sustaining cooperation thus may require a means to compel observance by raising the expected costs of violation – i.e. "enforcement".[1]

This can be a tall order given that a defining feature of the contemporary international system is the absence of a singular authority capable of policing adherence to international law, or of determining the allocation of penalties. Those seeking to promulgate robust international legal rules therefore often try to build monitoring and enforcement mechanisms into agreements. For complex multilateral agreements in particular, this commonly entails delegating these decisions to third party administrative, arbitral, or adjudicatory bodies.[2] By increasing transparency and lowering the political costs of penalizing violations, these bodies help to incentivize compliance in situations where concerns about reputation or ensuring the "rule of law" alone may not suffice.[3]

But, even if we allow that deterring non-compliance is the main reason states create third party tribunals, it does not follow that sanctioning blatant violations of legal rules will be their main activity in practice.[4] To the contrary, the more credible the enforcement threat posed by a tribunal the less it will be used for this purpose. What makes such a threat credible? Compulsory respondent

---

[1] The micro-foundations of this approach are traceable to widely accepted ideas about collective action problems in which the credible threat of costly punishment and the assumption of repeated interactions are key to finding cooperative equilibria. See Olson 1965, Axelrod 1986, Ordeshook 1986.

[2] Goldstein et al 2000. See also Milgrom, North, & Weingast 1990, Greif *et. al* 1994, Fearon 1998, Koremenos *et al* 2000, Simmons 2000, Smith 2000, Goldstein *et al* 2000, Rosendorff 2005, Barton *et al* 2006, Alter 2008 & 2014, Davis 2009 & 2012, Hartigan 2009:341, Satler, Spilker, and Bernauer 2013.

[3] Smith 2000, Abbott & Snidal 2000, Steinberg 2004, Davis 2009.

[4] Alter 2014 argues that international courts contribute to legalized governance in four ways: through dispute settlement, administrative review, enforcement, and constitutional review. However the existence of a dispute is common to all four functions, although their character varies. My argument is that the dynamics of enforcement differ systematically from these other three categories in ways that can be productively theorized.

participation, and a tribunal's ability to authorize remedies with "bite". This prompts a further question: if high capacity international tribunals are not calling out and penalizing opportunistic violations of known rules, *what are they doing*?

I argue that high capacity tribunals are more often preoccupied with resolving ambiguities in the applicability, scope, or content of legal rules. They do so by deciding disputes – but not just any disputes: highly concrete disputes over whether applicable law exists, or what it requires among parties with direct stakes in the outcome. They are, in short, *elaborating the law*.[5] My choice of verb here is considered. To "elaborate" entails adding detail to an extant rule or standard through interpretation. The added detail is in some sense "new", but it is also inextricably tied to the thing being interpreted.[6] Rule elaboration is an unavoidable externality of interpreting and applying ambiguous legal rules.[7]

Of course this is not to suggest enforcement-type disputes never involve ambiguities or uncertainties. Disputes alleging *prime facie* rule violations may encompass uncertainty about whether or not certain actions took place, or whether and to what degree harm was caused. There may likewise be uncertainty about whether a complaining state is sufficiently resolved to pursue a claim or to follow through on permitted remedies, or whether the respondent breached an obligation willfully or out of incapacity. The point, however, is that what is uncertain or ambiguous in these examples is something (or things) *other than the law*.

Why does underscoring this difference matter? I argue that enforcement and rule elaboration each involve distinct logics for drawing parties into legal disputes, and for selecting disputes into litigation. International disputes that begin with an enforcement profile have their fulcrum in conduct that is widely understood to be inconsistent with applicable legal rules. Here formalization of the dispute puts institutional pressure on violators by exposing rule-inconsistent behavior, or by validating the imposition of penalties. Enforcement actions also can have a socializing function.

---

[5] The applied character of deciding disputes among actual parties transforms interpretation from abstract argumentation to observable practice. This, in turn, contributes to its political relevance.

[6] As Dickson (2010) explains "interpretation is a Janus-faced concept, encompassing both a backward-looking conserving component, and a forward looking creating one" (citing Fiss 1982, Dworkin 1986, Marmor 1992, Endicott 1994, and Raz 1996).

[7] Rule elaboration is not a task of judicialized bodies only. It may encompass, instead, or in addition, the inputs of administrative bodies or the applied interpretations of agreement drafters and treaty secretariats (Alter 2014). It also may include the unprotested practices of states (especially powerful ones) where those practices are affirmative evidence of a customary or treaty rule (Putnam 2009 & 2016).

They can bring new treaty members, or states that have (for whatever reason) underinvested in learning about treaty rules, up to speed on the existence or character of their obligations.

This is in clear contrast to incentives underlying disputes with a rule elaboration profile. Here the core dynamic does not involve countering decisions to cheat. Instead, it is about grappling with ambiguities in whether or how legal rules apply. It follows that lumping contests over competing understandings of legal obligation together with efforts to call out willful violations under the label of "enforcement" may underwrite assumptions about the intentions and capabilities of disputing parties that are a poor fit for the interactions under study. (For example, dispute initiators may be cast as more virtuous and capable than is merited empirically, and respondents as more villainous or bumbling.) Consequently, defaulting to "enforcement" to describe legal disputing may systematically overstate the incidence of non-cooperative behavior. This can prompt illusory puzzles on the front end of research, and contribute to faulty inferences on the back end.

Below I theorize the distinction between rule elaboration and enforcement further and formulate expectations about when we should see recourse to third-party adjudication under each. Thereafter I illustrate the argument with data from WTO dispute settlement cases on subsidies. The results underscore that distinguishing between the rule elaboration and enforcement functions of international tribunals offers a more institutionally resonant view of international adjudicatory bodies, and a more richly strategic view of patterns of legal disputing and dispute resolution.

2.1 How is Enforcement Different than Rule Elaboration?

Game theorists have long recognized that the political world contains a variety of collective action problems, and that some are easier to solve than others.[8] On the easy end of the spectrum are "pure" coordination problems where the challenge involves establishing a common focal point among equally preferred options. Such situations are rare in international law, since few areas of international lawmaking involve blank legal and policy slates at the domestic level. Even so, convergence on a shared rule can be Pareto-improving even with adjustment costs if the bargain lasts beyond a certain period. But treaty members expecting to incur high short-term adjustment costs may not be eager to begin carrying out their obligations. The "shadow of enforcement" thus provides a needed nudge toward implementation.

[8] Olson 1971, Axelrod and Keohane 1985, Snidal 1985, Morrow 1994 and 2014.

Where incentives to cheat on agreements persist, this sets up a more difficult set of problems captured by the "Prisoner's Dilemma" game and its many variants.[9] In grappling with these issues, IR theorists have focused on identifying general incentives and disincentives for states to cooperate, along with the features of institutions that make those commitments credible (or not).[10] This literature has been illuminating in many respects. At the same time, it has largely bracketed questions about indeterminacies of law, treating legal rules as if the full extent of their applied meaning(s) is known to – or at least discoverable by – all relevant actors.[11] Indeed, the notion that regulated actors share common understandings (or conjectures) about the rules is at the heart of claims about how enforcement works, and why even its "shadow" can affect behavior.[12]

Taking a different tack more familiar to practicing lawyers, judges, and (gasp!) legal philosophers, I start with the idea that law is rife with ambiguities. This is an intuitively satisfying place to build a foundation of for examining the politics of legal contestation since, as an empirical matter, the scope and meaning of legal rules are rarely (if ever) wholly determinate, particularly at the margin.[13] Moreover, what rules mean may be highly actor and context dependent. Legal meanings are also subject to change over time. As such, making a legal commitment is far from the end of bargaining over the scope and content of legal rules. Instead, it is the start of a new and differently structured phase of contestation. On a textual level, limits of linguistic expression, coupled with the drafting generality of most treaties and statutes allow space for contestation. Efforts on the part of regulated actors to adapt law to changed conditions – whether generated by developments exogenous to the law, or from attempts on the part of strategic actors to manipulate the scope or meaning of their own or others' legal obligations – are another source of indeterminacy.

Of course, I am not arguing that IR studies of international dispute settlement ignore rule ambiguity altogether. Many empirical studies draw attention to "incomplete contracts" and problems generated by unanticipated "shocks". Judges are frequently described as engaging in "gap-filling," "rule clarification," or "judicial lawmaking"[14] However, these terms are rarely accompanied by

[9] Morrow 2014.
[10] Keohane 1984. Fearon 1998, Koremenos *et al* 2001.
[11] Hadfield and Weingast 2012 and Finnemore and Hollis 2016.
[12] Morrow 2014:282-283.
[13] Hart 1961.
[14] The existence of "gaps" may be the result of purposeful choices by treaty drafters in order to make it easier for states to reach bargains *ex ante*, or to allow members leeway to adopt different pathways to agreement objectives *ex post*. See Smith 2000, Abbott & Snidal 2000, Pistor & Xu 2002, Steinberg 2004.

concise definitions or focused theorization of the role of adjudicatory bodies in the production of legal meaning. This is both a cause and a consequence of the bracketing described above.

Consider, for example, that although a judicial decision may indeed "clarify" whether or how a law applies in a specific case,[15] this is not synonymous with boosting the clarity or precision of the rule at issue, or of rules overall.[16] First it is entirely possible for a legal rule to be applied in a way that underscores (or "clarifies") its permissive character, thereby undermining its precision. Second, litigated disputes often involve close questions of law or fact. They also may require judges to balance competing principles – such as whether health or environmental concerns may trump free market principles under a specific trade or investment law. Under such conditions slight variations – in the underlying facts, in legal framings, or in the training or temperament of the deciding judge(s) – can affect decisions. Thus even if we concede that judges generally aim for legal and institutional coherence, comparing judicial outcomes over time often exposes inconsistencies that confound rule clarity (or at least provide fuel for divergent interpretations on the part of decentralized actors).

Of course disputes may easily encompass elements of both rule elaboration and enforcement. A dispute with multiple claims may involve some features where the law is clear, and others where it is less so. In addition, dispute elements that begin as points of legal ambiguity can morph into situations requiring enforcement if the losing party is unwilling, or unable, to accept and implement the tribunal's decision. The focus of this inquiry is on differentiating the profile of claims on the front end of formal international disputing, since it is this configuration of beliefs and expectations that determines whether a dispute is formalized. However, enforcement challenges also may arise endogenously in the disputing process. I discuss this in section 2.4 after first expanding upon the nature of different informational barriers to negotiated solutions and their relation to selection into formalized disputing.

2.2 Rule following and incomplete information

Many rationalist treatments of strategic behavior incorporate some notion of "incomplete information" to describe gaps of knowledge about actor attributes or features of decision environments. This modeling heuristic comes with baggage, however, in the form of a strong

[15] Here the difference between "clarification" and "elaboration" may be negligible, though I prefer the latter for its semantic flexibility.
[16] Goldstein *et al.* 2000, 412-413 observe that precision matters because it "narrows the scope of reasonable interpretation."

supposition that missing information exists somewhere as discoverable fact.[17] In some instances this supposition may be reasonable. For example, where a state that invests few resources in acquiring (or maintaining) information about its international legal obligations finds itself in another government's legal crosshairs, this can prompt respondents to make that investment, to allocate more resources to compliance, or to reveal information about it unwillingness or (in)ability to do so.[18] From the other direction, initiating a dispute can demonstrate the ability of the complaining party to observe non-compliant behavior, or its resolve to insist that others follow the rules.[19] Critically, these information deficits each presume that parties agree, or can be compelled agree, that applicable law exists, together with a "correct" interpretation of its meaning as applied to the facts at hand. But what if the applicability of the rule or a shared interpretation cannot be presumed?

Taking a step back, where the behavior of interest concerns rule following, there are at least three generic dispute profiles encompassing differ types of information deficits. In the first everyone agrees a given practice violates the law, but they disagree over whether the practice is occurring, or on what scale. These are "fact-intensive" disputes. Second are situations in which parties acknowledge conduct, but disagree about whether it is prohibited.[20] These are "law-intensive" disputes. Because actual disputes can easily encompass both types of agreement, we also need a third "mixed" category. This category, in turn, can be subdivided into purely additive mixed disputes (characterized by distinct disagreements over fact and law), and those in which they are intermingled – as when the primary disagreement concerns the legal method by which key facts are established or how they are weighted (*i.e.* procedure).

Each of these profiles involves genuine disagreement. However, only the first ("fact intensive") type is an easy fit with standard rationalist assumptions, since only here can it be presumed that the information needed to resolve the dispute is "discoverable". The remaining types depend on two (or more) parties holding mutually incompatible beliefs about whether a legal rule exists, to whom it applies, or what it requires.[21] Where incompatible beliefs are sincerely held *and*

---

[17] Jupille *et al* 2013:34 distinguish between "risk" (which involves probabilistic calculations over known parameters), and "uncertainty" (which includes unknown possibilities and probability structures). Uncertainty, they insist, is not calculable using standard rationalist tools. Thus where risk modeling tools are used, it implies certain assumptions about the character of missing information—irrespective of whether those assumptions are descriptively valid.
[18] Chayes and Chayes 1992.
[19] Carnegie & Carson 2015, Kucik and Pelc 2016. Cite Trachtman!
[20] Note that here the dispute can be either about a characteristic of the rule, or about whether the rule applies at all.
[21] Putnam 2016. In many legal settings, the complainant must have, as a formal matter, suffered specific harm from allegedly prohibited acts, or from a failure to take required actions to have standing to raise a complaint. See the selection and scope conditions discussions below for additional criteria.

rooted in plausible theories of legal obligation, a solution cannot be "discovered" in the sense that it exists as a legal or social fact.

By implication, when a complaint features a law-intensive dimension, the parties' inability to reach a shared understanding may not be because one or both parties has under-invested in understanding the law, or its own or others' capacity to fulfill legal obligations. It may instead be because *an authoritative interpretation of the law in applied form does not (yet) exist.*[22] Where legal rules are not wholly determinate, no side need be acting in bad faith, or be wildly misinformed, for disputes to arise. Furthermore, under these conditions, legal deterrence cannot operate – except to limit the range of plausible beliefs regarding the existence of legal indeterminacies.[23]

Once a tribunal interprets and applies a previously indeterminate rule, its decision is a legal and social fact. This changes the game vis-à-vis the parties, such that subsequent compliance (or non-compliance) reflects most direct on the adjudicatory authority of the tribunal. Whether and how individual decisions affect the larger *tableau* of prior practice is a further question.[24] Here too caution is needed, since the broader applicability of any decision may be hotly contested by other observers of judicial rulings. [25]

2.3 Selecting into adjudication

Understanding why standard assumptions of the enforcement frame do a poor job of describing important subsets of adjudicated disputes requires attention to characteristics of dispute formation and selection into litigation.[26] For a legal dispute to arise at least one party must believe it has been harmed by the failure of another to follow applicable law, and the accused party must disagree – either about the permissibility of the behavior, or about the extent of the harm. For a dispute to reach litigation, however, each side must believe, in addition, that it is likely enough to have its own view vindicated by a competent tribunal to justify the costs and risks of litigating.[27] Operative thresholds for this type of risk thus turn on what is at stake for each party in having its

[22] Raz 2009. Legal lacunae of this type may arise anywhere within legal regimes, but are especially likely with new rules, or where rules are being stretched to cover new actors, issues, or technologies, Putnam 2016.
[23] Likewise, the incapacity story that is central to the managerial school of international compliance is logically coherent only where it is relatively clear what legal rules require.
[24] If a judicial decision does not purport expressly to overrule a prior decision, or decisions authoritatively, I assume both continue to be valid. This is particularly important in international law, where norms of formalized precedent are weak.
[25] Note, even in precedent-based legal systems, outcomes that are 'limited to their facts' exist. This is generally due to use of equity principles, i.e. where straightforward application of a legal rule will yield an unfair or absurd result.
[26] Putnam 2009 & 2016.
[27] For this to hold, litigation must not be more costly or less effective than available alternatives, Putnam 2016.

interpretation of the law vindicated. In litigation in international courts, the nature and strength of state preferences for particular interpretations of legal rules generally turn on domestic politics.

Where prospective complainants lack sufficient confidence in their interpretation of the law, their own stakes are not especially high, or there is reason to doubt that an authoritative statement of the law will help to change the respondent's behavior, disputes will not be formalized. Further, where respondents recognize that the law's requirements are already clear, or that their stakes in the dispute are low, they rationally will opt to settle by reaching a "mutually agreed solution" with the complaining party.[28] In situations where a state's domestic law on a contested issue is new, or not well developed, adjustments may be easier than for established rules or entitlements.

Instances requiring "enforcement" can still arise when violators lose bets on the quality of monitoring, or on others' willingness to overlook rule deviations – "slack" in the system.[29] Also, some fact-intensive disputes may be amenable to settlement only after formal proceedings begin. This may occur, for example, where the consultation (or investigation) process provides respondents with new information about underlying facts, about the law, or about the complainant's resolve. The remaining law-intensive and "mixed" disputes should be those in which parties genuinely disagree about some aspect of the law or its associated procedures, and the outcome matters to both parties. Stated more technically, litigation selects for situations where the applicability, boundaries, or content of legal rules are *not* common knowledge in a game-theoretic sense.[30]

An important exception concerns instances in which states have large stakes in getting to a preferred interpretation of a rule, but the chances of prevailing in the tribunal in question are slim. This is most likely for disputes that implicate arrangements with deep roots in a respondent's domestic legal and institutional arrangements, or which will negatively affect politically powerful domestic constituencies.[31] Under these conditions, international adjudication offers a chance to

[28] Priest and Klein 1984, Putnam 2009 & 2016, Alschner 2014. However, Shavell (1996) underscores that this probability may be unequal if underlying stakes are unequal. Note, however, that international litigation generally involves only highly sophisticated actors. As such, the stakes may encompass not only those of the dispute at hand, but also estimates of how the legal rule is likely to be interpreted in the future under all possible outcomes, see Putnam 2009 & 2015, Pelc 2012, Galanter 1974.

[29] Acknowledging that such slack exists, however, does not negate arguments that suggest that developing countries have been less able to use the DSU in sophisticated ways as compared to larger and legally more well-resourced trading partners. Both phenomena can be occurring simultaneously.

[30] Putnam 2016, Priest and Klein 1984.

[31] Davis 2000, Satler *et al* 2013, Putnam 2016.

secure a preferred legal interpretation [32]—or, failing that, a source of political cover.[33] In the meantime, the disputing process itself "buys time" for respondents to continue the behavior at issue.

Law has (and is intended to have) distributional effects: it is a medium for influencing who gets what, when, and how. We might expect regulated actors to generally prefer versions of rules that favor their own interests and goals (all else equal). In many situations there will be room for more than one good faith (if instrumentally convenient) interpretation of a legal rule.[34] Where divergent interpretations of legal rules clash, international tribunals serve to mediate the ability of more powerful parties to dictate what international commitments mean, or when they apply, by claiming authority over such determinations.[35]

Insisting upon analytic space for good faith divergences over the scope or meaning of legal rules in no way suggests governments and private entities behave in an other-regarding manner when they champion specific legal interpretations, or when they take issue with those of other regulated actors. To the contrary, I expect states and sub-state actors to bet regularly on how much (or little) rules can be stretched to fit their own preferences before generating pushback.[36] Nor do I suggest that states will never try to fudge the boundary between good faith and bad faith challenges to the scope or meaning of commitments. However, even allowing that all states may sometimes act out of duplicity or incapacity,[37] I expect shared awareness that self-serving behavior may occur around legal interpretations, in turn, to cast its own shadow over interactions among regulated actors. Specifically, it should shape how governments attempt to justify their behavior, and limits their ability to adopt idiosyncratic interpretations of the law in order to unilaterally alter the terms of international bargains.[38]

Taking a broader view, the incidence of claims and their selection into litigation should vary cross-sectionally by legal domain and by tribunal character. According to a 2004 comprehensive survey by the Project on International Courts and Tribunals, the most "judicialized" areas of

---

[32] Pelc 2012.
[33] Allee and Huth 2006, Bown 2009, Creamer 2015. Sykes argues instances of "efficient breach" should be common. Pelc (2010) finds empirically they are not.
[34] This simultaneity can exist because few laws involve wholly unconditional requirements or prohibitions, which allows for context-dependence in how rules are interpreted and applied. This malleability extends to how broadly or narrowly specific disputes are framed, and how contested categories are defined, cf. Guzman & Simmons 2002.
[35] Venkze 2012.
[36] As others have observed, complainants in WTO disputes are structurally inclined to favor measures to increase trade openness. In general, this involves efforts to compel other states to eliminate policies and practices that are alleged to impede trade open trade (Tarullo 2003, Barton *et al* 2006).
[37] Simmons 2000, Neumeyer 2005, Chayes and Chayes 1993.
[38] Stated in more technical terms, DSBs place a curb on auto-interpretation in international law.

international affairs by a considerable margin are economic and commercial matters, and international human rights.[39] Even so, the existence of multiple tribunals alone says little about their individual capacities, or about underlying patterns of use.

Three factors are especially important here. The first is standing – the legal capacity to initiate claims. The more inclusive the rules of standing, the greater the potential that a tribunal will be used, *ceteris paribus*. Next is the capacity of the court to compel the participation of respondents. The technical term for this is "compulsory jurisdiction". Third, is a belief on the part of prospective litigants that losing parties will accept outcomes and follow through on remedies. Where complaints are infrequent, respondent non-participation is common, or judgments are frequently ignored or fudged, I expect international dockets to become thinner and weighted toward tackling clear instances of violation.

2.4 Distinguishing types of non-compliance underlying enforcement

It is also important to distinguish between challenges to the scope or content of legal rules as formulated by regulated actors, and challenges to the *authority of institutions* tasked with interpreting and applying those rules. Whereas the definition of enforcement used above refers to the first of these, when lawyers and judges refer to "enforcement" they typically mean the second. The key idea here is that once a tribunal has ruled on a matter, it is, by definition, no longer ambiguous – at least vis-à-vis the disputing parties. The relevance of the decision to other parties and other disputes – its "precedential value" – may be contested legitimately. But, these two types of challenges are not equivalent. When recipients of adverse judgments defy a tribunal's orders, this threatens not just the integrity of the rule, but also to the authority of the deciding tribunal.

Where resistance to a tribunal decision is observed, it is possible that it was there from the start. Or, it may have emerged endogenously from the process of disputing. For example, a respondent that is surprised to have its interpretation of a rule rejected, or heavily qualified may choose to dig in and resist implementing the tribunal's recommendations. It some instances resistance might be limited to a specific decision or an issue-specific cluster of decisions – as with Israel's pointed defiance of the International Court of Justice's ruling on the international legality of its boundary fence with the Palestinian territories. In others it may be indicative of broader defiance

[39] Romano 2004.

the structure or objectives of a legal or policy regime, as with Russia's pattern of disregarding decisions handed down by the European Court of Human Rights.

A final (but related) issue of relevance is time. Should we expect the distribution of enforcement versus rule clarification to be roughly constant over the life of an institution, including following large legal or organizational changes? Or should it vary—and why? A view of dispute settlement grounded in enforcement logic might lead one to expect lots of claims early in an institution's lifetime (*i.e.* as the DSB's competence is probed and challenged) but few settlements. It would also suggest the rate of fully appealed cases would be higher early on, or immediately after substantial reorganizations, but to decline thereafter.[40] Increases in the complexity of claims would be attributed to disputing parties becoming more sophisticated about how to use dispute settlement bodies strategically.[41]

By contrast, the rule elaboration logic would expect that in the period shortly after a tribunal with compulsory jurisdiction is established, enforcement claims and simple coordination-type claims over category boundaries and institutional procedures will be relatively more likely. A high proportion of these claims should settle, except when adjustment costs for the respondent are anticipated to be high, in which case the dispute should be resolved after an initial ruling. Later, as the legal framework matures, and the "low hanging fruit" (in terms of easy to resolve disputes) is plucked, more legally complex issues will arise as tensions or inconsistencies across various aspects of the agreement come to light, and as members wrestle with new and unforeseen developments. Legally complex claims (meaning those with greater numbers of plausible tie-ins to other treaty elements) will also be more likely to be litigated through the appellate stage.

A slight twist on this argument concerns new treaty members. New members generally have less information about the rules and procedures of the regime than more long-standing members. Late joiners, unlike original members, thus encounter a set of potentially non-obvious, but actually discoverable (as in having been previously authoritatively decided) interpretations of legal rules and procedures. Consequently, we might expect relatively high rates of enforcement claims against new treaty members. This serves to inform new members about (or socialize them to) the organization's

[40] Thus, in this initial period, a mechanism either will prove itself unhelpful and states will not expend scarce resources to use it, or initial questions concerning its jurisdiction and authority will be settled, and thereafter disputes will mainly concern issues of substance or remedies.
[41] Bown 2004, Bermeo & Davis 2008, Pelc 2012.

workings. This includes learning about other members' tolerance for rule stretching, and on foot-dragging on implementing shared obligations.

In summary, if we accept that adjudication selects for law-intensive disputes with high domestic stakes for the parties involved, political considerations suggest these cases will be not only less likely to settle, but also more likely to be appealed. When parties to these types of disputes are unsuccessful, the same firm, industry and sectorial pressures that led to participation in international litigation can prompt governments to delay adjusting in order to buy time for additional violations, to minimize the adjustments they make, or even to pay assigned penalties for continuing to violate.[42]

Disputing in the WTO: An Empirical Proof of Concept

Testing the intuitions developed above requires differentiating empirically between rule elaboration and enforcement. This is most tractable at a relatively micro level in high capacity tribunals where legal disputes arise regularly. I focus here on the World Trade Organization (WTO) and its Dispute Settlement Body (DSB). Within this still sizeable area of activity, look to disputes over state subsidies (state actions that have the intent or effect of conferring trade advantages on specific firms or sectors) and countervailing measures. There are several reasons for these choices.

First is the WTO's centrality to international trade. A foundational IPE claim is that international trade is rife with incentives to cheat, since individually states are better off if the rest of the world removes trade barriers when they do not.[43] In a world of imperfect monitoring and enforcement, therefore, it is reasonable to expect some behavior that is clearly contrary to the letter and the spirit of trade liberalization – even among states that value open trade as a public good. The WTO's DSB is widely credited with helping to smooth expansion of the WTO regime, albeit in ways that have generally favored the interests of more powerful members.[44] Second, international trade is an area where incentives for states (individually) to violate rules are often strong. As such advances in trade liberalization are not simply a function of states acting on prior preferences. Third, the WTO is a global organization. Consequently, DSB outputs generated through the behavior of its members cannot be explained as an artifact of any particular country or region.

---

[42] Brewster 2009. This may be especially true at the WTO, where the *ad hoc* character of the panels may encourage appeals if the losing side views its loss as an artifact of the panel's composition.
[43] Goldstein and Krasner 1984, Gowa 1995, Rosendorff and Milner 2001.
[44] Critics of the WTO are many and varied. They include states and sub-state actors that disagree with the objectives of the WTO and with its means. Some commentators view the DSU system as part of the problem.

Other reasons for focusing on the WTO relate to research design. The institutional change at the heart of this examination—the adoption of the 1995 WTO Agreements as a "package deal"—affected all members simultaneously, making it a good test bed for the argument from a research design perspective. Also, the fact that WTO dispute settlement has several well-defined stages allows for observing when disputes are resolved (or abandoned). The choice to focus on a single trade issue is for tractability. Although WTO disputes constitute a relatively manageable universe of cases in some respects, I suspect that a convincing initial effort at differentiating enforcement profiles from those involving legal ambiguities may require attention to interplay between general characteristics of disputing, and issue-specific detail. At the same time, subsidies and countervailing measures claims occur together with other claimed violations of other WTO Agreement provisions within a single complaint. This allows for examination of whether and how different types of claims are tied in with other issues and broad principles. Finally, this analysis is an exercise in concept development. As such, beginning with a subset of WTO disputes reserves a substantial amount of international trade data for later theory building and testing.

Adapting my selection argument to the WTO context, I expect few disputes to begin with efforts to "enforce" established rules of the GATT/WTO in the face of flagrant violations. To the contrary, I expect the actual work of the DSB to most often involve resolving ambiguities over nominally plausible interpretations of ambiguous provisions in applied situations. Why? Because where the DSB's enforcement capacity is credible, the shadow of DSB proceedings should deter blatant violations of the WTO rules, and prompt quick settlements from such violators of this type identified by affected members.[45] If correct, this theoretical claim should manifest itself in observable patterns of WTO disputing.

The task of differentiating dispute profiles is complicated by the fact that individual disputes can have elements of both rule elaboration and enforcement. Sequencing matters here. Where disputes begin with a rule elaboration profile, enforcement problems may emerge endogenously. However, that doesn't change the essential character of the original dispute for classification purposes. The converse also can occur, where a respondent's sincerely held belief about the law is wholly unreasonable on its face, thus transforming a dispute that, from the respondent's perspective, was initially about legal ambiguity into enforcement. For reasons discussed above, I expect these

[45] The screening of states into WTO membership offers added insurance that incapacity will not be a common issue, see Von Stein 2005, Hopkins & Simmons 2005.

situations to be rare in the WTO context, and to arise mainly in situations involving new or under-socialized members of the regime.[46]

In the first stage of WTO dispute settlement, a state that claims harm from the policies or programs of another member requests consultations toward finding a mutually agreed solution.[47] If the parties fail to reach agreement, the complainant may ask the DSB to establish an *ad hoc* panel of three trade experts to investigate and decide the issue(s). Prior to 1989, GATT dispute resolution procedures required all parties to agree to panel submission, which gave defending states an effective veto over proceedings. Reforms adopted in 1989 gave complainants a "right" to a panel -- though the losing party retained the option to not adopt the resulting decision.[48]

The Dispute Settlement Understanding adopted as part of the 1994 Marrakech Agreement establishing the GATT/WTO further altered the regime's dispute settlement architecture by including a provision for the automatic adoption of Panel decisions, except where there is consensus *not* to adopt. It also created a seven-member standing Appellate Body (AB). Parties now may adopt Panel decisions "as is", or appeal specific points of law for appellate review.[49] Once the AB decides there is no further basis for appeal. Thereafter, parties subject to adverse rulings are expected to adjust their laws and policies accordingly, and to notify the DSB.[50] Lastly, if a winning party is dissatisfied with the pace or quality of implementation, it may request arbitration. If this fails to resolve the matter, it may ask the DSB to form a compliance panel, which can, if appropriate, authorize remedial sanctions.[51]

The choice to focus on disputes over subsides and CVDs was not motivated by prior knowledge that could bias the results. The objective was to focus on an area with a long history of regulation (unlike trade in services, phytosanitary rules, and intellectual property rights – all of which have a shallower provenance) to allow for comparison of institutional design effects. Under the GATT, subsidies were regulated on the basis of an *a la carte* agreement ("code"), as was typical of many pre-1994 non-tariff arrangements.[52] Like many elements of the trade regime, subsides rules

---

[46] In other words, these are situations with a prior case *directly* on point – *i.e.* where precedent applies straightforwardly.
[47] Hudec (1980) explains that dispute settlement became a GATT objective only in the Tokyo Round of negotiations.
[48] This set of arrangements gave rise to the infamous "Shrimp/Turtle Dispute" between the United States and Mexico.
[49] Although the AB is, in general, precluded from revisiting factual claims, a "back door" mechanism exists in Article 11 by which the AB may be asked to determine whether the Panel's assessment of the facts was "objective" – which may necessitate re-looking at the facts. My thanks to Jeffrey Dunoff for bringing this to my attention.
[50] DSU Article 3(6).
[51] Bernauer, Elsig, and Pauwelyn 2012, 486.
[52] Barton *et al.* 2006.

were heavily reworked in the Uruguay Round negotiations. Under the WTO, subsidies are covered by a non-optional Agreement on Subsidies and Counter-Measures (SCM) with its own Standing Committee.[53] Note that the SCM is long and technical relative to other WTO agreements. This might prompt expectations that underlying demand for rule elaboration might be less than for more generally drafted parts of WTO rules. Or, the abundance of technical detail might itself provide fodder for contestation over the scope and meaning of particular provisions.[54]

Subsidies rules implicate aspects of states' industrial policies and domestic welfare programs. These policies may fall under the WTO's purview when they begin to affect international terms of trade. This can happen when policies are aimed at making specific exports more attractive externally, or at making imports less attractive. Practices considered strictly off-limits include payments (rebates, loans, tax credits, or other benefits) that are contingent upon export performance, or which include local content requirements. Outside this category of prohibited subsidies, are "actionable" subsides, which require the complainant to demonstrate specific harm to its interests.[55]

A second type of dispute under the SCM concerns countervailing duties (CVDs). Under the SCM, if a WTO member finds a trading partner is using non-permitted subsidies, it can unilaterally impose measures to counteract the effects of the subsidy. These measures can then become the object of disputes where the targeted country alleges inadequacies in the imposing country's procedures for identifying subsides, or it finds the CVD improper or disproportionate.

Data and Results

The dispute data are from the WTO with an initial "leg up" from collection by Horn, Johannesson, and Mavroidis (2011) and Johns and Pelc (2015). As of October 2016, 513 complaints had been initiated at the WTO. Of these, 112 had a subsidies and countermeasures dimension.[56] I limit the analysis to disputes initiated through 2011, in order to ensure individual disputes have ample time to make it through the full AB process with standard margins of delay. This yields a universe of 89 cases and 455 separate subsidy and countermeasure claims. Not all of these cases

[53] The first part concerns definitions, prohibitions, and provisions for remedies (SCM Articles 1-10), along with provisions on procedural and evidentiary standards, and for the assessment, imposition, and suspension of countervailing duties (Articles 11-24). The final section has transitional provisions and measures for integration into the GATT and other preexisting arrangements.
[54] My thanks to Jeffrey Dunoff for highlighting this issue.
[55] Trebilcock 2011; Sykes 2002.
[56] This and subsequent descriptive statistics are drawn from data provided on http://www.worldtradelaw.net.

involve unique facts; a few are largely duplicative claims filed by different states against the same defendant.[57] However, because I am chiefly interested in the initial profile of disputes that select into litigation, I count each case with a unique dispute settlement number as an observation.[58]

Dependent variable

The dependent variable (DV) is the profile of disputes at the point of formalization. Unfortunately, the ability to observe only complaints that become requests for consultations precludes direct analysis of why certain disputes ripen into complaints and others do not.[59] While not ideal, these data nonetheless allow me to test for key implications of my theory – namely, that few enforcement cases (as defined here) will appear in the sample, and those that do will involve high respondent stakes, or respondents that are new WTO members.

As such, this analysis must be able to differentiate these disputes from those driven initially by disagreements over the scope or meaning of legal rules. To do so, I use computerized text analysis to construct a classification mechanism.[60] The approach, called "bag of words", de-structures documents into individual words and word strings for analysis of their frequency and distribution on grounds that this can reveal underlying patterns and structures not readily observable in the original documents.[61] The operative conjecture here is that disputes driven at the initial stage by efforts to call out clear violations of known rules should have textual elements that are distinct from disputes animated by legal indeterminacies.

Drawing upon the theory, I focus on two general categories of words to distinguish dispute types. First are words that indicate a belief on the part of complainants or adjudicators that respondents have acted in bad faith ("scolding" words). Second are words that reference the international trade regime and its objectives, on grounds that behavior that blatantly violates known rules is more likely to prompt these types of references ("system maintenance" words). I am also interested in terms that might be common only in disputes over indeterminacies in the law.

---

[57] For example, see disputes filed against Canada by Japan (DS 139) and the European Community (DS 142) concerning subsidies related to domestic auto manufacturing.
[58] Of the initial 89 cases, eight were ultimately merged, yielding a total of 81outcomes.
[59] Christina Davis is compiling a dataset of annual National Trade Estimates from trade ministries that will help to address this shortcoming by tracking which actions in which foreign countries were being flagged as potential violations of WTO rules. Similarly, Global Trade Alert crowd-sources measures that could lead to interstate trade disputes, although from 2009 onward only, http://www.globaltradealert.org/site-statistics.
[60] My thanks to Matthew Carr for extensive research assistance on this aspect of the project.
[61] Torkkola 2001. The technique is commonly used in the tech world for tasks like email "spam" filtering.

To specify lexicon of such expressions, I first looked to a set of non-subsidies cases widely known to involve overt violations of WTO rules: those concerning use of a method to calculate foreign underpricing of imports in anti-dumping disputes called "zeroing". The method was held to be WTO-incompatible in 2003.[62] However, the United States continued to use zeroing until 2012, prompting 14 additional complaints. For each, I gathered all documents on the WTO's website up through AB reports (if applicable). I excluded documents from within-dispute compliance proceedings, since these concern the second face of enforcement discussed above.[63] The segmentation yielded 9,549 unique words ("unigrams"), and 111,940 word pairs ("bigrams").[64] This de-structuring was repeated for the subsidies and countermeasures disputes using the same document exclusion and hit count protocols. This yielded 28,186 unigrams and 644,120 bigrams.

[TABLE 1 HERE]

I then manually identified terms that fit the above criteria, excluding generic terms (e.g. "volume", "export", "export volume"), terms specific to zeroing or dumping, and any references to particular entities or products. Finally, for each list, I discontinued the manual search once the count of hits was fewer than ten.[65] Because the subsides and countermeasures pool has roughly seven times more disputes than the zeroing pool, simply comparing raw hit counts would bias against finding enforcement cases for subsidies (and in favor of my theory). To correct for this, each n-gram count was recalculated as a ratio of the total number of words in each corpus. The difference between these ratios for each n-gram provides a normalized "score" of disproportionate usage in one corpus over the other. To further ensure the terms selected are representative of a particular class, I included only n-grams at or above the 95th percentile of over-representation in the zeroing cases ("enforcement terms"), and also a few with a similarly disproportionate incidence in the sample of subsidy and CVD disputes ("ambiguity terms"). This yielded an "enforcement dictionary" of 22 unigrams and 14 bigrams.[66]

[FIGURE 1 HERE]

[62] See DS141 (European Communities – Anti-Dumping Duties on Imports of Cotton-type Bed Linen from India).
[63] As a sorting protocol, I classified any document that references Article 21 of the WTO Agreement as involving a dispute over compliance with a WTO adjudicator specifically.
[64] Note, the program automatically excludes all articles, pronouns, and prepositions. This count, however, includes variants of words. Thus, "fact" and "facts" is two hits.
[65] This cut the sample to 4,782 unigrams, 14,854 bigrams.
[66] If a term was used in a selected bigram, its components were excluded from the unigram list to avoid double counting.

Next, the corpus of subsidy and CVD documents was searched for hits on the dictionary terms. Deciding what to draw from the yield is tricky. Raw hit counts are potentially misleading, since the longer a dispute goes on, the more words are produced. However, tracking the proportion of enforcement terms in each case may introduce a different bias – by discounting the early enforcement elements of complaints that become legally complex disputes. Furthermore, informal spot-checking showed the rate of false hits around several of the unigrams to be quite high, although not for the bigrams.[67] Subtracting the proportion of "indeterminacy" terms in each complaint from the proportion of enforcement terms helps to correct for each of these issues. As Figure 1 shows, it also produces a clear separation in the cases.[68]

Explanatory variables

For other variables, I record the year each dispute was formalized, whether it concerned subsidies or CVDs, and the stage at which it terminated. Prior work demonstrates that states are not all equally likely to use third-party tribunals -- whether for enforcement, or in efforts to shape international rules.[69] I thus track the power configuration between the main parties in each dispute in the sample using dummies for parties at rough economic parity as measured by 1996 GDP, for strong states formalizing complaints against weaker states, and for weaker states formalizing complaints against strong states. Where there is more than one entity in any position, I code for the most powerful in the set. Of the 89 subsidies and CVD complaints filed prior to January 2012, only 13 did not involve the United States or the European Community as primary parties.[70]

[FIGURE 2 HERE]

Busch and Reinhardt find a strong negative relationship between complexity and a complainant prevailing at the panel or AB phase (*i.e.* having most or all elements decided in its favor).[71] They ascribe this to variance in legal experience and sophistication. To address this, I count the number of SCM Agreement articles cited in each complaint, along with the total number of WTO Agreement elements (*e.g.* GATT articles, anti-dumping rules). As measure of WTO

[67] This aspect of the analysis will be more thoroughly checked using a Python-based extraction of every dictionary term hit in the data sample, along with its surrounding text.
[68] My thanks to Matt Carr for this strategy.
[69] Horn, Johannesson, Mavroidis 2011.
[70] Davis and Bormeo (2009) attribute the relative underuse of the WTO's DSB by developing states, and it's non-use by LCDs in part to high start-up costs due to a lack of direct experience with international level proceedings.
[71] Busch & Reinhardt 2002.

experience, I coded the length of parties' GATT/WTO membership, along with counts of complaints filed previously or defended (overall and re subsidies).[72]

Figure 3 lists each of the 32 articles of the SCM Agreement with bars indicating the number of times each is raised in the dispute sample. Provisions on the core prohibitions of the SCM are, not surprisingly, most common. These are closely followed by citation to Articles 10 and 23, which cover how the SCM Agreement relates to other WTO agreements (Other Final Provisions). It suggests further that a non-trivial amount of WTO disputing concerns neither enforcement nor clarification of substantive rules but *procedure*.

[FIGURE 3 HERE]

Without disagreeing that prior experience may matter, I argue, in addition, that disputes that require balancing different principles are likely also to be complex—albeit in different sense. For example, cases that are primarily about establishing boundaries on accepted obligations or refining procedural rules should have a comparatively simple structure (citing at most a handful of provisions with little effort to draw in other agreements). From the opposite end of the spectrum, it may be that complaints involving deep disagreements over how to balance persistent inequities in WTO law will be "complex" in the sense measured here – meaning they may reference a wide array of Agreement provisions, including provisions spanning other (non-subsidy) elements of the regime. As such, my priors on this point are mixed.

Prior studies of WTO disputing have identified the importance of controlling for differential domestic stakes.[73] However, the empirical tests on which these findings rely often use fairly chunky categorical variables.[74] I try to improve upon this by dividing each party's annual international trade in goods into four industry-based bins (agriculture, energy, manufactures, and mining). I then divide each bin by the country's total exports to a get a rough measure of stakes. Industry data are from the World Development Indicators dataset.[75] Data on annual totals of exports by country are from COW National Trade 3.0 from 1996-2009, supplemented by data from the IMF's International

[72] Busch & Reinhardt 2006. I also gathered data tracking state participation as a third party in subsidies-related disputes, although I don't (yet) use it in the analysis.
[73] Guzman and Simmons 2002, Davis 2008, Sattler *et. al.* 2013.
[74] A partial exception is Busch & Reinhardt 2006:464, which uses the log of product-specific respondent imports with dummies to distinguish non-goods-centered disputes. However, since importers are often direct beneficiaries of foreign subsidies, for my purposes is it more appropriate to focus on the export implications for exporters and importers on the logic that the import price distortions of foreign subsidies harm the ability of domestic industries to compete.
[75] My thanks to Joonseok Yang for sharing his collected data standardized to ISIC2.

Trade Yearbook for 2010 and 2011. As an alternative, I include a dummy variable for agricultural disputes.[76]

Finally, I include POLITY IV measures for regime type, since Bown (2004) and others find democracies are systematically more likely to engage in WTO disputing.

Descriptive results

Figure 4 shows the distribution of the outcomes for the 89 subsidies and countermeasures cases (SCM) terminated at the WTO between 1996 and 2011. The elements of the figure are consistent with the theory, with the highest concentration of pre-panel settlements coming near the beginning of the period of study. Figure 4(a) indicates the large number of consultations that did not proceed to the Panel stage, or to a formally notified solution. Some of these complaints may have been abandoned due to reassessments on the part of the initiator. Others may have been resolved by respondents discontinuing the targeted conduct.

[FIGURE 4 HERE]

Figure 4(b) shows that pre-panel settlement as defined here was, as expected, most prevalent immediately after the WTO's creation. In line with the socialization aspect of the theory, it also shows an uptick in pre-panel settlements in 2006 following the expiry of China's "unofficial grace period" for adjustment following its 2001 accession to the WTO.[77] The relatively large number of complaints in 2004 (3 of 6) reflects a spike in consultations over CVD procedures.[78] Also as expected, 4(c) shows that the number of SCM disputes terminated after Panel proceedings (only) is small, and skewed toward the middle. Of the eleven cases in this group, three were flagged as "enforcement" by the classification analysis. Two were part of the long-running softwood lumber dispute between the United States and Canada – and thus did not truly terminate at this stage. Of the remaining six, five involved less powerful states going after more powerful states unsuccessfully – meaning the Panel found the respondent largely not in violation.[79]

The last element of Figure 4 contains 31 disputes that reached the Appellate Body (AB). The raw numbers here are higher than anticipated in the early part of the WTO era, although they are a

[76] Busch & Reinhardt 2006:463 also track "politically sensitive" disputes, which are those where respondents try to justify conduct on non-commercial grounds – for e.g. environmental or health protection, or national security. Only one such dispute exists in the subsidies sample, so I do not include control for it here.
[77] Zeng & Liang 2013:242.
[78] See DS 310, DS311, and DS 314.
[79] The final case, DS154 (Indonesia – Autos) concerned whether certain local content requirements fell within temporary developing country exceptions. The Panel held they did not.

relatively modest proportion of all complaints initiated in this period. Seven of these disputes were flagged as "enforcement". Three are subsidies claims – two (against Canada, and United States) alleging aircraft industry subsidies, and another (also against Canada) alleging that premiums given to companies for producing renewable energy locally, thus excluding non-local producers, constituted a "specific" subsidy.[80] Each activity arguably falls within core SCM Article 3 prohibitions. The remaining flagged disputes involve CVDs. Three allege *ad hoc* decision making that was either inconsistent with the respondent's own national legislation (Japan – CV Duties on DRAMS), or that relies on faulty national legislation (Mexico – Anti-Dumping Measures on Beef and Rice, China – CV and Anti-Dumping Duties on Steel). The fourth is a dispute between the United States and the EC over valuation of a one-time pre-sale benefit to a public entity when determining whether a post-privatization owner is receiving a subsidy (with the EC maintaining any benefit would have been absorbed in the sale price). All allege fairly egregious (from a WTO perspective) deficiencies either in the respondent's domestic laws, or with some aspect of their application.

Figure 5 breaks down complaints in the sample according to the power configuration of the primary disputants, and the stage at which the complaint terminated. The y-axis in each of the subfigures is the stage reached, with "0" indicating pre-panel settlement, "1" indicating termination after a Panel report, and "2" indicating termination after an Appellate Body decision. Rows (1), (2), and (3) show data from complaints between the United States and the European Union, complaints requested by more powerful states against less powerful states, and by less powerful states against more powerful states, respectively.[81] The dispute dots reflect variation in the number of separate claims in the complaint, such that additional claims increase the circle size, and have been jittered to improve legibility. The left column of Figure 5 shows SCM references only. The right column shows the total number of claims in each complaint.

[FIGURE 5 HERE]

Row 1 shows complaints involving entities roughly balanced in economic power. The relative sparseness of dispute dots near the "0" on the y-axis indicates little early settlement activity, and no terminations occurred after the Panel phase only. Rather, when disputes involving these two

[80] http://www.reuters.com/article/us-wto-idUSBRE9450HA20130506.

[81] For this exercise, I look only to the dyadic comparison between the main state parties to a consultation request in making a more/less powerful making—not to broader categories of "developed", "developing", etc. Also, where a WTO complaint names a specific European country as the respondent, and not the European Union, I code the power configuration as stated—meaning I do not assign the dispute to the EU.

entities are formalized, appellate review is often required. This likely reflects an especially high capacity for leveraging potentially beneficial indeterminacies in the law, along with the existence of more avenues for resolving "easy" cases without litigation.

By comparison, Row 2 shows that early on, and again after 2005, powerful states raised a barrage of complaints against weaker states that were not pursued to the Panel stage. As the theory expects, complaints reaching the AB are more common over time. Row 3, which tracks complaints by less powerful states against more powerful states, shows a more evenly dispersed trend of complaints not reaching the Panel stage, and few early complaints reaching the AB. One curious feature of the complaints initiated by weaker states that terminated after the panel stage is that they are tightly clustered in the period between 2000 and 2005.

In Figure 5, the expected negative correlation between the number of claims per complaint and the incidence of pre-panel settlement is evident across the board, but especially with respect to disputes involving asymmetric parties. This confirms prior work. More to the point of this paper, Figure 4 also indicates claims have gotten more complex over time. Under the logic of "enforcement" this would imply that violations have grown more egregious as the institution has matured. My theory provides an alternative explanation – namely, that in the early years of WTO a relatively high proportion of requests involved longstanding complaints that would have been blocked by the respondent under the old rules, along with coordination-type issues tied to questions about the procedure and competences of the institution itself. The modal profile of complaints thus should shift over time toward more-difficult-to-resolve issues of principle balancing and equity – except where new states join.

Another plausible alternative is that poorer states, and those less well integrated into the global trading system, *are simply less happy with WTO rules*, particularly those involving arrangements (like rich-state agricultural subsidies) that restrict their ability to compete. Thus, when WTO members choose to take on an economically more powerful state on broad issues of equity, their claims will tend, by necessity, toward the complex, requiring efforts to draw from foundational WTO principles and analogous areas of law and practice. And, because DSB actions (particularly those of the AB) more closely mirror the preferences of the economically powerful, those states have few incentives to concede early.[82]

[82] Barton *et al* 2006:60.

With the above in mind, I next turned to regression analysis to test for additional associations among the variables described. In light of the rareness of one of the two possible outcomes of the DV and the relatively small sample, I opted to use a penalized maximum likelihood estimator "firthlogit" (similar to exlogit). The method aims to overcome "separation" in logistic regression, described as "a condition in the data in which maximum likelihood estimates tend to infinity (become inestimable)."[83]

[TABLE 2 HERE]

The results of the regression analysis add little to the descriptive exercises. Running the model restacking variables and alternative specifications yields reliably statistically significant coefficients for *YEAR_INITIATED* (at the 99% level) only, along with coefficients significant at the 90% level for *TOTAL_NUMBER_CLAIMS*. Both of which have a small positive relationship with enforcement-type disputes. Using only the number of subsidies claims to capture "complexity" yielded similar results, but measuring the proportion of subsidies claims fell well short of any significance. Coefficients for the categorical elements of *POWER_CONFIGURATION* do not reliably predict enforcement-type claims. Moreover, the most substantively suggestive results from this cluster indicate that weaker states are especially unlikely to formalize these complaints, contra my theoretical expectations. The export-based measures of complainant and respondent *STAKES* used separately and as a difference function likewise indicate little consistent effect. Finally, neither the Polity IV scores of complainants and respondents analyzed separately, nor the difference between complainant and respondent scores help to predict enforcement-type complaints.[84]

Discussion

Drawing attention the rule elaboration function of adjudicatory bodies sheds new light on some old puzzles in the international trade dispute literature, and likewise gives rise to a number of fresh lines of inquiry. I now highlight two. The first involves changes over time in the character and composition of formalized international trade disputes. The second concerns implications for repeat litigation and precedent.

In examining WTO disputes, Busch and Reinhardt (2003) find, strikingly, that neither the design modifications of the GATT dispute resolution mechanisms in 1989 nor the WTO's Dispute

[83] For detail see Heinze and Schempter 2002, Coveney 2008.
[84] I was unable to check the effects of the categorical variables in the model reported in Table 2 using simulation-embedded first differences, since Clarify (Tomz *et al* 2006) does not support firthlogit.

Settlement Understanding appeared to have affected rates of disputing, or rates of settlement for filed complaints.[85] For scholars accustomed to expecting institutional design to matter, this is deeply puzzling. One possibility suggested by the analysis here is that even if *rates* of disputing and settlement have not changed markedly, the *composition of complaints* has changed. Embracing the distinction between enforcement and elaboration offers a window onto how the logics of pre- and post-1995 disputing differ.

Whereas the GATT dispute mechanism was helpful in getting past what in other contexts might be called "mutually hurting stalemates", disputing at the WTO has, in addition, the capacity to discipline states that are clearly benefiting from behavior that violates WTO rules. This change has enhanced enforcement muscle of the institution, and arguably also altered the behavior of its members. That is to say, state behavior that would by untouchable under a "right-to-panel but non-automatic-adoption" regime may be suppressed at the source under the shadow of more robust enforcement. Conversely, under pre-1995 arrangements requiring Panel decisions to be adopted by consensus, states had few incentives to file complaints involving far-reaching or highly contentious issues.[86] In sum, the compulsory character of WTO dispute settlement, coupled with largely automatic adoption of panel and AB decisions, has enabled an entirely new class of contentious, law-intensive issues to see the light of day.[87]

A second point of focus in recent work on international adjudication concerns expectations for repeat litigation of similar claims (*i.e.* the effects of precedent).[88] Under a pure enforcement view, it is entirely possible to see largely identical disputes being filed again and again, since, as argued above, the primary challenge to be managed is one of slack.[89] The theory of rule elaboration articulated here, however, suggests we should rarely, if ever, see identical disputes being litigated through the panel stage, since the law in question already exists in authoritative form. Exceptions may arise when offending states lack capacity to comply with the rules, where they are trying to test the capacity of the enforcing institution, or where they are engaging in "efficient breach" -- when states knowingly allow violations because the expected cost of penalty is less than the cost of

[85] Busch and Reinhardt 2003:145.
[86] Following Alter (2014), I expect this logic to apply only for compulsory jurisdiction. Although states do sometimes submit legal disputes to third parties on a voluntary basis, the conditions under which this occurs are limited.[86]
[87] Alter 2014:278-279.
[88] Pelc 2014, Verdier and Voeten 2015.
[89] That said, among states concerned with the normative costs of being caught in illegal behavior, a high proportion of such claims will settle, Busch & Reinhardt 2002.

discontinuing the violation.[90] This argument is thus compatible with those that highlight more strategic uses of disputing in efforts to refashion international legal rules.[91]

Each of these situations is more likely where domestic stakes are high and/or powerful domestic lobbies make it difficult for governments to alter existing laws and entitlements to bring them into line with international rules.[92] Under these circumstances additional violations may accrue, prompting compliance proceedings and/or follow-on disputes. Insistence that repeat claims are unlikely does not, however, preclude *similar disputes* from arising as states seek to push or probe the limits of prior rulings.[93] This is in part a function of the endogeneity of disputing: when courts and other judicialized tribunals issue decisions, this may resolve some legal questions, but give rise to others as the implications of each new ruling percolate through the legal framework as a whole.[94]

This suggests interactions leading to the formalization of disputes may vary depending on whether powerful states are on the complainant or the respondent side – or both. For example, we might expect that where underlying rules are clear, stronger states will prefer to avoid the cost, delays, crowding out other complaints associated with formally requesting consultations. (Indeed, the SCM expressly provides a mechanism for unilateral action to address impermissible subsidies via CVDs.) However, weaker states might turn to the DSB where rules are clear as a means to demonstrate resolve.[95]

That said, the enforcement frame is more appropriate to some legal issue areas than others. For example, in the spheres of international human rights and international criminal law, a key challenge has been to create and sustain institutions capable of providing any enforcement at all. As such, use of criminal tribunals is at present generally limited to allegations of fairly egregious substantive violations. Here too, however, there is often plenty of action around clarifying jurisdiction in individual situations, adjudicating fairness of rules of evidence and other procedures, assessing the permissibility of various theories of liability, and grappling with issues of remedies.

---

[90] Posner and Sykes 2011, Morrison 1994. Here DSBs are mechanisms for allocating side payments. Pelc (2010) argues such behavior should be rare, since governments (unlike trade economists) do not view violations in isolation. Factoring in broader costs of allowing some states to "punch and pay" to regime support makes efficient breach far less efficient.
[91] Pelc (2012), for example, is careful to argue that only a subset of WTO disputes involve test case-type claims, and that this is a practice that is generally limited to economically more powerful states.
[92] Davis 2012.
[93] Putnam 2015. Pelc 2013 goes so far as to suggest that especially sophisticated WTO members may purposeful engineer multistage lines of disputes calculated to reach a preferred interpretation of particular rules.
[94] Alter 2003, Putnam 2015.
[95] Bown 2004, Davis and Bermeo 2009. Here I refer to weakness strictly in relative terms within dispute dyads.

The theory and data analyzed here suggest that the profile of complaints filed at the WTO should change, and has changed, over time as the institution has matured and as new members have become socialized to WTO norms and practices.

Conclusion

Drawing a conceptual distinction between rule enforcement and rule elaboration matters for how we describe, theorize, and model behaviors that feed international dispute settlement, and for the inferences we draw about the effects – and effectiveness – of international tribunals. The tendency to describe third party adjudication as "enforcement" risks overstating the incidence of blatant forms of cheating in the international system, and imposes an overly punitive frame on the function of law in guiding behavior.[96] To the contrary, much of what the WTO's DSB, and other high capacity institutions, do involves interpreting the applicability, contours, or meaning of ambiguous agreement provisions in concrete settings—*i.e.* rule elaboration.[97]

Theorizing the existence and character of the rule interpretation and elaboration function of DSBs opens up a rich and largely unexplored universe of strategic behavior.[98] Much of the IR literature on commitment keeping fuses these concepts and then labels the entire package "enforcement". Behavior that is determined to deviate from authoritative interpretation of the rules is then attributed either to actions taken in bad faith, or to legal or institutional incapacity to comply. The possibility that parties might reasonably disagree about what a rule means or how it applies is often not expressly considered.

This paper argues, to the contrary, that state parties select themselves into litigation most often precisely when they insist on different and incompatible interpretations of legal rules and procedures. Approaching these disputes as "enforcement" problems is thus a categorical error since it presupposes something that does not (yet) exist. This paper aims to provide a bridge to more theoretically informed ideas about when actors are likely to use third-party DSBs for rule-clarification, and when they are likely to use them for enforcement—both generally, and in the context of particular institutions. At present, it clearly raises more questions than it answers, but further iterations will aim to shift this balance to address at least the meatiest of them in light of

[96] Chayes and Chayes 1992.
[97] Orakhelashvili 2008:2.
[98] See, however, Pelc 2010, Johns and Pelc 2015 who argue that WTO litigants attempt to use litigation strategically—albeit, without theorizing the rule elaboration role of DSBs.

dominant claims in the literatures on WTO dispute resolution, and processes of international disputing more generally.

What might be said more broadly? As has been observed, the WTO's DSB authority is unusually robust due to its compulsory jurisdiction over members backed by high stakes for states to remaining within the regime.[99] This makes the WTO a necessary case for my claims, but also a limiting one. Indeed, I expect that when judicialized institutions are less robust (meaning less able to compel the participation of respondents, or to authorize remedies), that prospective litigants will bring fewer claims, adjusting for membership and transaction volumes. Furthermore, the claims that are brought will involve either relatively blatant violations of known rules, or low stakes coordination issues. This conjecture is testable both across various international institutions, and inter-temporally with respect to pre-WTO GATT dispute procedures.

This paper does not grapple with the empirics of wins and losses at the DSB. Such questions, however, are related to the main focus here: stage of termination. When dealing with questions of enforcement, the rate or success for complainants (and here I regard favorable settlements as successes) should be very high, especially when controlling for power and the intensity of preferences as measured by the domestic political dominance of the affected industries. By contrast, rates of litigation should be low. This is so, since, by definition, in such instances the main barrier to compliance in enforcement claims involves a failure of will or capacity to comply – not ambiguity about what is required.

With complaints that are primarily about resolving ambiguities in the law, by contrast, we should expect rates of litigation to be much higher, but rates of complainant success to be lower. Precise predictions regarding how much lower the complainant success rate should be, however, are not possible absent information about the stakes for the involved parties.[100] Still, given the ability of complainant states to choose the claims they want to pursue, one wouldn't expect the aggregate success rate to dip much below 50 percent as long as states have some alternative (non-WTO) political and legal options for influencing other states.[101]

To bring the argument full circle, collective decisions to create robust institutions for international dispute settlement indeed may be chiefly about deterring cheating. Even so, effective

[99] Alter 2014.
[100] Shavell 1996.
[101] Priest and Klein 1984, Putnam 2015.

enforcement requires deterring not only blatant violations, but also dis-incentivizing opportunistic behavior in fuzzier areas. As such, the tendency to fuse enforcement and rule elaboration understates the degree to which law and legal contestation may serve as an adaptive medium in the face of new and unanticipated developments.

The implications for international politics and for rational design are substantial: allocating interpretative authority to adjudicatory bodies in settings where participation is compulsory—as is the case in an ever-expanding roster of international institutions—means judges and other similarly empowered authorities will, in some sense, "make law".[102] This triggers meaty normative and quasi-constitutional questions about whether judicial functions can, or should, be strictly limited to "applying" the law as written, or whether they encompass authority to adapt legal rules to changing circumstances so as to fulfill their "object and purpose". However, this dialectic of opportunistic tinkering on the part of domestic actors that stand to gain from particular interpretations of legal rules, and pushback from those who claim to be harmed by those interpretations, is what propels processes of legal adaptation—domestically, transnationally, and internationally.[103]

Acknowledging the rule-generative elements of legal interpretation should not, however, prompt the types of fears sometimes expressed about over-reaching and unaccountable international adjudicators. Courts are, by and large, highly constrained institutions.[104] Their authority is delegated by political processes, and is at all times highly mediated by jurisdictional rules. Perhaps most critically, judicial tribunals have no direct enforcement power. Instead, they rely on others to respond to their pronouncements. International courts, with few exceptions, are limited to deciding only questions that relate specifically to the treaty to which they owe their existence. Without denying that this type of institutional setup can underwrite a number of political and jurisprudential pathologies, the threat of global judicial dominance is not one of them.

[102] Alter 2014.
[103] Putnam 2016.
[104] See Steinberg 2004.

WORKS REFERENCED

Abbott, Kenneth W, Robert O Keohane, Andrew Moravcsik, Anne-Marie Slaughter, and Duncan Snidal. "The Concept of Legalization." *Power and Governance in a Partially Globalized World* (2000): 132-51.

Allee, Todd L., and Paul K. Huth. 2006. "Legitimizing dispute settlement: International legal rulings as domestic political cover." *American Political Science Review* 100:2, 219-234.

Alschner, Wolfgang. 2014. "Amicable settlements of WTO disputes: bilateral solutions in a multilateral system." *World Trade Review* 13:1, 65-102.

Alter, Karen J. "Resolving or Exacerbating Disputes? The WTO's New Dispute Resolution System." *International Affairs Int Affairs* (2003): 783-800.

Alter, Karen J. "Delegating to International Courts: Self-Binding vs. Other-Binding Delegation." *Law and Contemporary Problems* 22 Mar. 2008.

Alter, Karen J. *The New Terrain of International Law: Courts, Politics, Rights*. Princeton: Princeton UP, 2014.

Barbieri, Katherine and Omar Keshk. 2012. Correlates of War Project Trade Data Set Codebook, Version 3.0. Online: http://correlatesofwar.org

Barton, John H., Judith L. Goldstein, Timothy E. Josling, and Richard H. Steinberg. 2006. *The Evolution of the Trade Regime: Politics, Law, and Economic of the GATT and WTO*, Princeton University Press.

Bernauer, Thomas, Joost Pauwelyn, and Manfred Elsig. 2012. "Dispute Settlement Mechanism—Analysis And Problems." *Oxford Handbooks Online*.

Bown, Chad P. 2009. "US-China Trade Conflicts and the Future of the WTO. "*Fletcher Forum for World Affairs*. 33, 27-48.

Bown, Chad P. 2005. "Participation in WTO Dispute Settlement: Complainants, Interested Parties and Free Riders." *The World Bank Economic Review* 19, 287-310.

Broude, Tomer. 2004. *International Governance in the WTO: Judicial Boundaries and Political Capitulation*, Cameron May.

Brewster, Rachel. 2009. "Shadow Unilateralism: Enforcing International Trade Law at the WTO." *University of Pennsylvania Journal of International Economic Law* .

Busch, Marc L., and Eric Reinhardt. 2006. "Three's a Crowd: Third Parties and WTO Dispute Settlement." *World Politics* 446-77.

Busch, Marc L., and Eric Reinhardt. 2003. "The Evolution of GATT/WTO Dispute Settlement." *Trade Policy Research*.

Chayes, Abram, and Antonia Handler Chayes. "On Compliance." *International Law and International Relations* (1993): 65-91.

Davis, Christina L., and Sarah Blodgett Bermeo. "Who Files? Developing Country Participation in GATT/WTO Adjudication." *Jop The Journal of Politics* (2009): 1033-049.

Davis, Christina L. *Why Adjudicate? Enforcing Trade Rules in the WTO*. Princeton: Princeton UP, 2012.

Downs, George W., David M. Rocke, and Peter N. Barsoom. "Is The Good News About Compliance Good News About Cooperation?" *International Organization* (1996): 379-406.

Dworkin, Ronald. 1986. *Law's Empire*. Fontana Press.

Endicott, Timothy A.O. 1994. "Putting Interpretation in It's Place," *Law and Philosophy* 13, 451-479.

Fearon, James D. "Bargaining, Enforcement, and International Cooperation." *International Organization* (1998): 269-305. Print.

Gaubatz, Kurt Taylor. (1994).

Goldstein, Judith, and Richard Steinberg. "Negotiate or Litigate? Effects of WTO Judicial

Delegation on U.S. Trade Politics." *Law and Contemporary Problems* (2008): 257-82.
Gowa, Joanne. *Allies, Adversaries, and International Trade*. Princeton: Princeton UP, 1995.
Greif, Avner, Paul Milgrom, and Barry R. Weingast. "Coordination, Commitment, And Enforcement: The Case Of The Merchant Guild." *Journal of Political Economy* (1994).
Guzman, Andrew, and Beth A. Simmons. 2002. "To Settle or Empanel-An Empirical Analysis of Litigation and Settlement at the World Trade Organization." *J. Legal Stud. S205* 31.
Hadfield, Gillian, and Barry R. Weingast. "Microfoundations of the Rule of Law." *SSRN Electronic Journal SSRN Journal* (2013).
Heinze, Georg, and Michael Schemper. 2002. "A solution to the problem of separation in logistic regression." *Statistics in medicine* 21:16, 2409-2419.
Hopkins, Daniel J., and Beth A. Simmons. "The Constraining Power of International Treaties: Theory and Methods." *Am. Pol. Sci. Rev. American Political Science Review* (2005).
Horn, Henrik, Louise Johannesson, and Petros C. Mavroidis. 2011. "The WTO Dispute Settlement System 1995-2010: Some Descriptive Statistics." *SSRN Electronic Journal SSRN Journal.*
Johns, Leslie, and Krzysztof Pelc. "Fear of Crowds in WTO Disputes: Why Don't More Countries Participate?" *SSRN Electronic Journal SSRN Journal* (2013).
Koremenos, Barbara, Charles Lipson, and Duncan Snidal. "The Rational Design of International Institutions." *The Rational Design of International Institutions* (2001): 1-40.
Marmor, A. 1992. *Interpretation and Legal Theory*. Clarendon Press.
Mnookin, Robert H., and Lewis Kornhauser. "Bargaining in the Shadow of the Law: The Case of Divorce." *The Yale Law Journal* (1979).
Olson, Mancur. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard UP, 1965.
Orakhelashvili, Alexander. *The Interpretation of Acts and Rules in Public International Law*. Oxford: Oxford UP, 2008.
Pelc, Krzysztof J. "Constraining Coercion? Legitimacy and Its Role in U.S. Trade Policy, 1975–2000." *International Organization Int Org* (2010).
Pelc, Krzysztof J. "The Politics of Precedent in International Law: A Social Network Application." *American Political Science Review* (2014): 547-64.
Priest, George L., and Benjamin Klein. "The Selection of Disputes for Litigation." *The Journal of Legal Studies* (1984).
Putnam, Tonya L. (2016). *Extraterritoriality, Domestic Courts, and the International Politics of Jurisdiction* New York: Cambridge University Press (forthcoming).
Raz, Joseph. 2009. *Between Authority and Interpretation*, Oxford University Press.
Raz, Joseph. 1996. "Why Interpret?" *Ratio Juris* 9, 349-63.
Rosendorff, B. Peter. 2005. "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure" *American Political Science Review* 99:3, 389-400.
Rosendorff, B. Peter and Helen Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape" *International Organization* 55:4, 829-857.
Sattler, Thomas, Gabriele Spilker, and Thomas Bernauer. 2014. "Does WTO dispute settlement enforce or inform?." *British Journal of Political Science* 44,:04, 877-902
Sebastien, Thomas. 2007. "World Trade Organization Remedies and the Assessment of Proportionality: Equivalence and Appropriateness" Harvard International Law Journal 48:2, 337-382.
Shavell, Steven. 1996. "Any Frequency of Plaintiff Victory is Possible" *Journal of Legal Studies* 25, 493-502.
Smith, James McCall. 2000. "The Politics of Dispute Settlement Design: Explaining Legalism in Regional Trade Pacts" *International Organization* 54:1, 137-180.

Steinberg, Richard. 2004. "Judicial Lawmaking at the WTO: Discursive, Constitutional, and Political Constraints" *AJIL* 98:2, 247-275.
Sykes, Alan O. 2003. "The Economics of WTO Rule on Subsidies and Countervailing Measures" John M. Olin Law and Economic Working Paper No. 186 (2d Series).
Torkkola, Kari. 2001. "Linear discriminant analysis in document classification." In *IEEE ICDM Workshop on Text Mining*, pp. 800-806.
Trebilcock, Michael J. 2011. *Understanding Trade Law* Northampton, Mass: Edward Elgar Publishers.
Verdier, Pierre-Hugues, and Erik Voeten. 2014. "Precedent, compliance, and change in customary international law: An explanatory theory." *American Journal of International Law* 108:3, 389-434.
Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance" *APSR* 99:4, 611-622.
Zeng, Ka and Wei Liang. 2013. *China and Global Trade Governance: China's First Decade in the World Trade Organization* New York: Routledge.

Table 1: Bigrams in Term Search Dictionary

Scolding Words	*System Maintenance Words*	*Indeterminacy Words*
Absurd unreasonable	International obligation	Good faith
Ambiguous obscure	Multilateral trading	Ordinary meaning
Arbitrary unjustified	Purposes WTO	Plain meaning
Inherently unfair		
Legally flawed		
Plain language		
Prima facie		
Without merit		

Table 2: Results of Regression Analysis Using Firthlogit

Total Claims	.08*
	(.04)
Subsidy or CVD	-.36
	(1.06)
Strong vs Weak	-.26
	(1.07)
Weak vs Strong	-1.58
	(1.06)
Difference in Stakes	-1.51
	(1.24)
Difference in Polity IV	-.07
	(.06)
Year Formalized	.1***
	(.02)
Constant	-198.16****
	(38.27)
N	75
Chi2	0.0

Figure 1: Classification Results

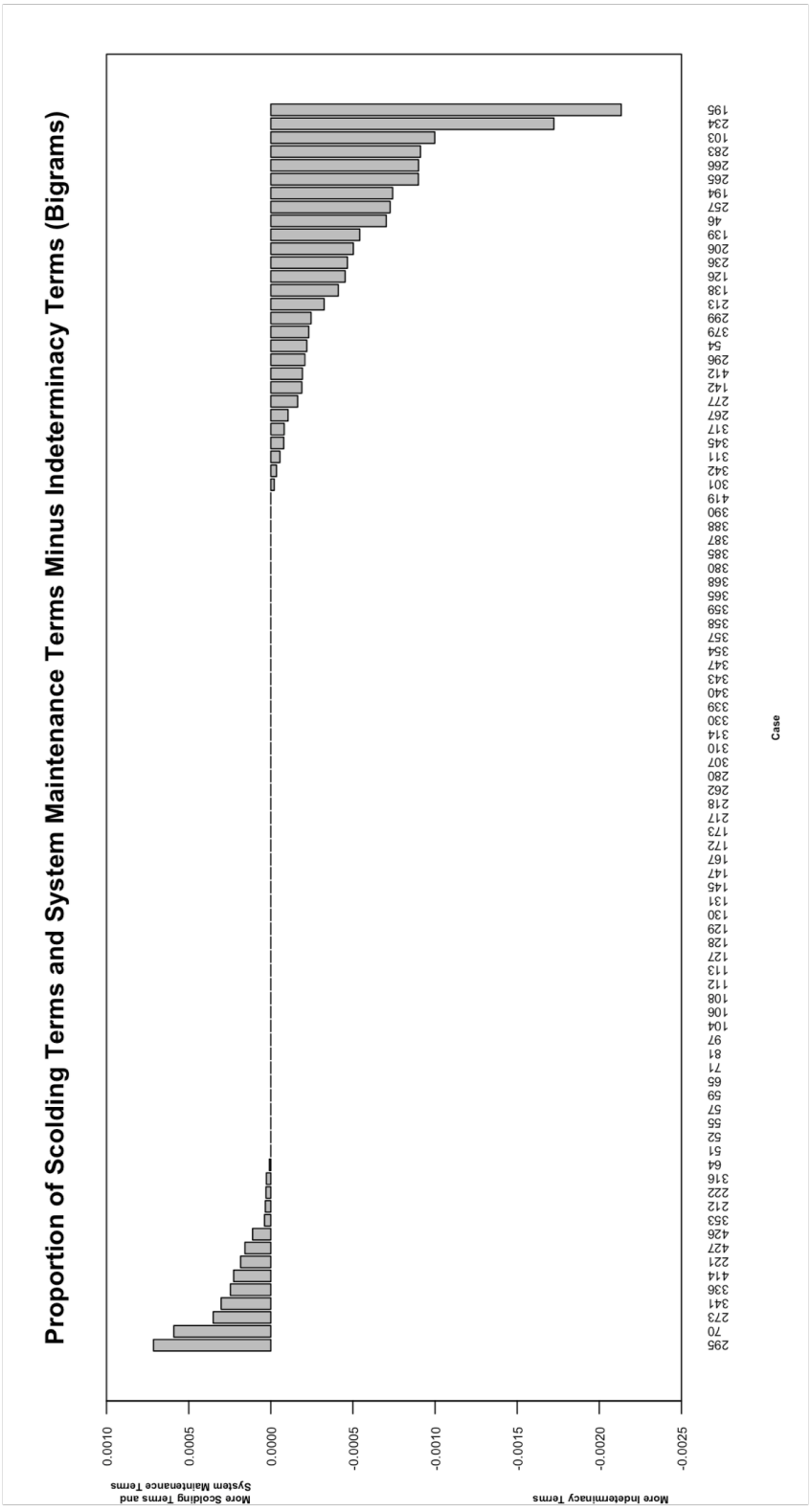


Figure 2: Subsides and CVD Complaints by Year Initiated

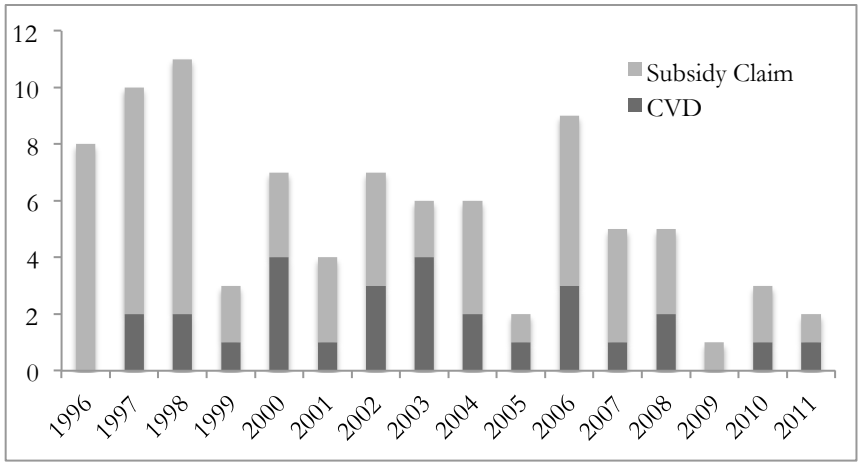
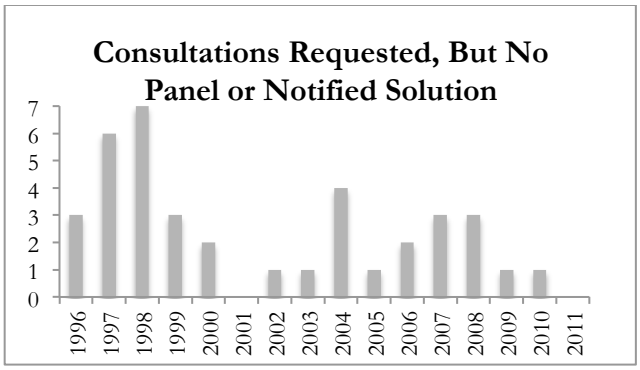
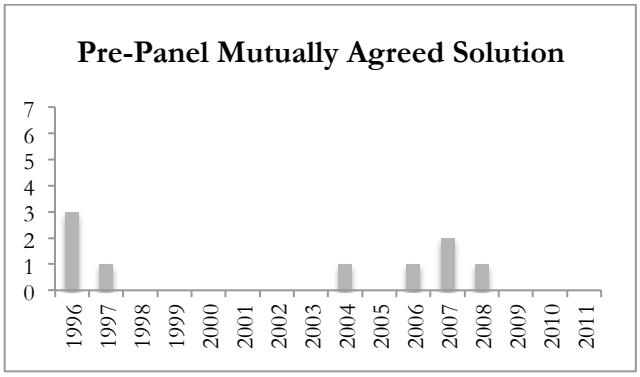


Figure 3: Distribution of Terminations in WTO Disputes over Subsidies and Countervailing Measures

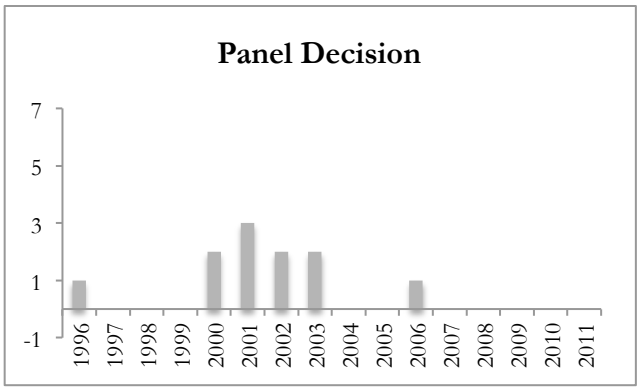
(a)



(b)



(c)



(d)

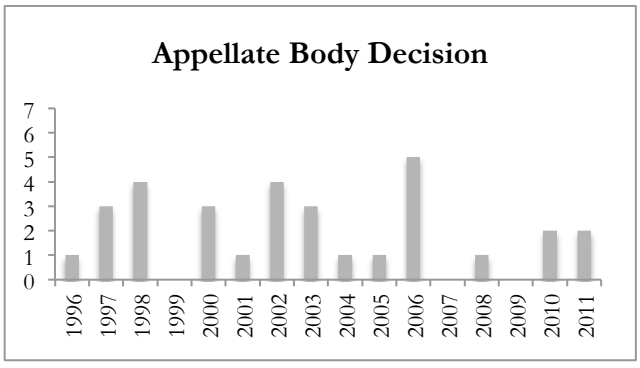
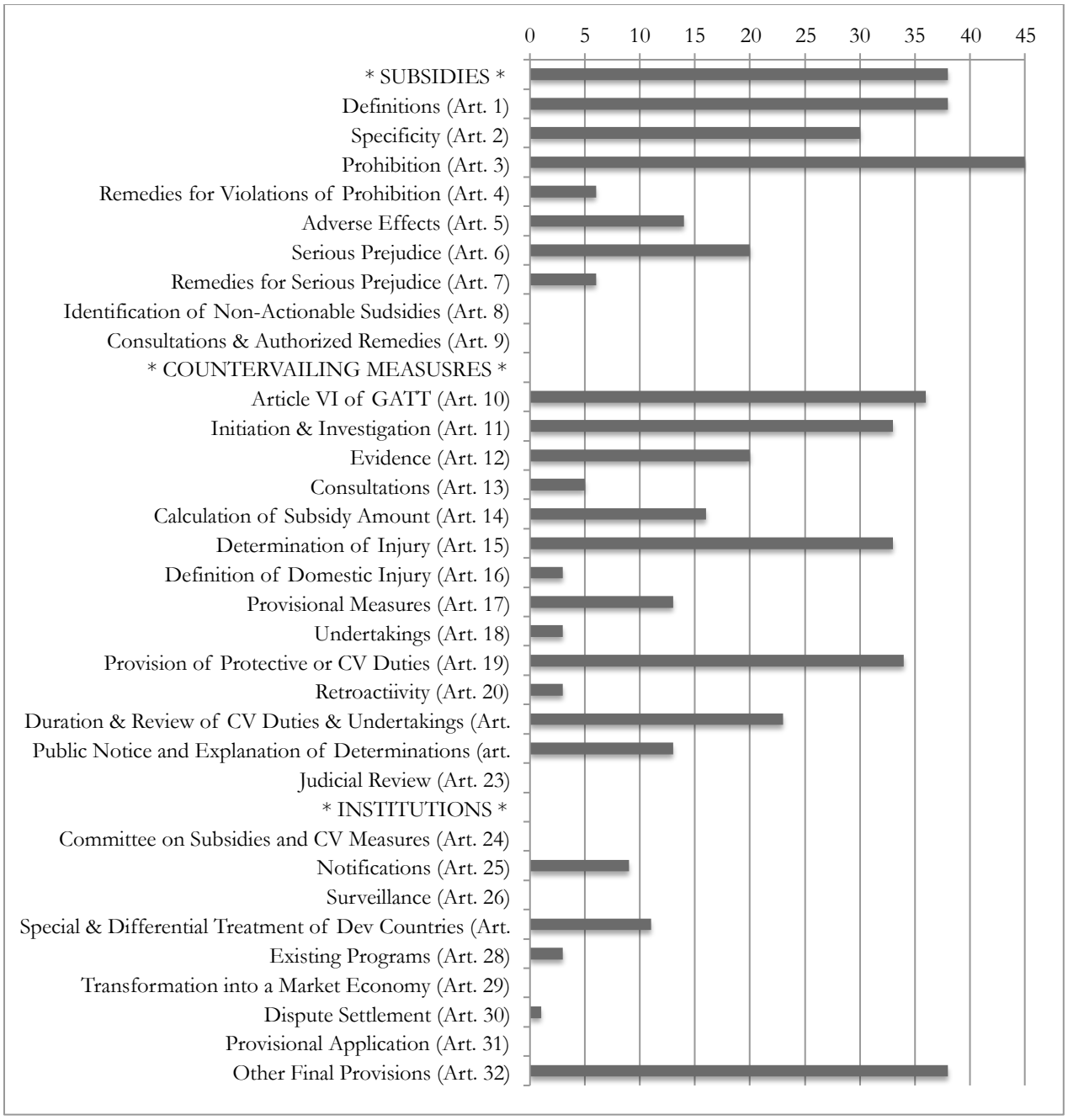


Figure 4: Count of SCM Articles Cited in WTO Consultation Requests 1996-2011 (N=261)[105]



[105] Note, some complaints contain more than one citation to each article, which is not reflected here.

Figure 5: Stage Reached by Year, Power Configuration of Main Parties, and Weighted by Number of Claims