Free Riding on Enforcement in the World Trade Organization

Leslie Johns, University of California, Los Angeles
Krzysztof J. Pelc, McGill University

Many policies that appear to violate WTO rules remain unchallenged, even as they have a significant economic impact. Why is this? We argue that the likelihood that a country challenges a protectionist policy is linked to how concentrated or diffuse that policy is. When it is concentrated, litigation is a private good. But when a policy is diffuse, affecting many states, litigation is a public good and countries face a collective action problem: each country seeks to free ride on others' litigation. The resulting selection effect has two consequences. First, we see a longer enforcement delay for diffuse trade violations. Second, states require higher odds of success to overcome the collective action problem, meaning that conditional on being filed, cases that challenge concentrated policies are less likely to succeed. Examining all WTO disputes, we leverage selection effects to test our argument using data on the timing and outcomes of trade disputes.

Like many international organizations, the World Trade Organization (WTO) relies upon its members to challenge possible violations. At times, its dispute settlement system appears highly effective, with members challenging protectionist trade policies as soon as they appear. For example, the United States blocked entry to Canadian trucks carrying cattle and swine in 1998, citing health concerns. Canada believed that this policy, which affected only Canadian trucks, violated WTO rules. Canada responded quickly: 15 days after the policy was implemented, Canada filed a dispute and requested expedited consultations with the United States.[1]

Yet the WTO's dispute settlement system sometimes appears ineffective, with trade violations going unchallenged for years. For example, the US Agriculture Improvement and Reform Act of 1996 (FAIR Act) violated WTO rules by subsidizing US corn exports. Because subsidies depress world prices and corn is a widely produced commodity, the FAIR Act harmed a great number of countries. However, Canada was the only country that ultimately paid the cost of challenging the FAIR Act, and it did so 4,025 days—over 11 years—after the law's implementation.[2] Why did Canada swiftly challenge the United States in 1998 over a relatively limited trade restriction on cattle and swine, while ignoring large US corn subsidies for over a decade? More generally, why are there greater enforcement delays for some policies than others?

Our claim is that Canada's decision can be explained in part by how the US trade policies affected other countries. Trade policies vary in both their aggregate economic impact, as measured by the amount of trade at stake, and the distribution of that impact across states. Some protectionist policies have relatively small but highly concentrated effects, like the US measure on Canadian trucks. Others have far larger but more diffuse effects, like the US corn subsidies. Regardless of the aggregate economic impact of trade policies, variation in their diffuseness determines the extent to which challenges of these policies represent public goods.

We argue that policy diffuseness has two overlooked effects, which we derive using a formal model. First, we argue that holding the aggregate economic impact of a trade policy

1. DS (dispute settlement) 144: *US—Certain Measures Affecting the Import of Cattle, Swine and Grain from Canada*.
2. DS 357: *US—Agricultural Subsidies*.

constant, a country should wait longer, on average, to file a case against a more diffuse policy. So conditional on a policy being challenged at the WTO, a more diffuse policy will have been in place longer than a more concentrated policy. Second, we examine the predicted legal success of WTO cases. Some potential cases are more likely to generate a procomplainant ruling, because the underlying facts and applicable law are more favorable. A country must carefully weigh a potential case's legal merit against its cost. For a concentrated policy, a country will only file if the case has a sufficiently high chance of success. For a diffuse policy, a country has an even higher standard: given the temptation to free ride on others, a potential case must have an especially high chance of success for an affected country to file it. We do not directly test which cases are selected for litigation; instead, we derive predictions about the likely success of cases that are selected and test this against the data. We show that in observed WTO disputes, cases that challenge concentrated policies succeed less often than cases that challenge diffuse policies. So the temptation to free ride produces enforcement delays and results in the wrong cases being filed.

We test our arguments on 360 WTO disputes from 1995 to 2013. To measure enforcement delay, we use data from Bown and Reynolds (2015) on the implementation dates for all policies challenged at the WTO. To test our predictions about success in litigation, we construct an original data set of the ruling direction—whether the complainant prevailed on a legal argument—for all 4,484 individual claims brought before the WTO during the period of interest.

FREE RIDING ON ENFORCEMENT

The crown jewel at the core of the contemporary international trade regime is the WTO's dispute settlement understanding (DSU). This international legal body is built on decentralized enforcement: while it has no centralized prosecutorial function, it allows countries to challenge one another's policies, and in the absence of a mutually agreeable settlement between the two parties, it adjudicates the matter and offers recommendations for compliance. These are binding in nature. In its scope, its rate of compliance, its rich jurisprudence, the WTO's legal body is arguably the most successful court in the international system. Nonetheless, there are limits on its effectiveness. Namely, the WTO can only adjudicate disputes that states bring before it. As in all legal systems, not all violations that occur in the trade regime are challenged. While the exact proportion is difficult to estimate, most trade lawyers would agree that most violations are never challenged.

Enforcement is costly, and the government agencies that oversee WTO disputes lack the resources to challenge all possible violations. Even the United States and the European Union are severely constrained by underfunding. Average annual spending by the Office of the US Trade Representative (USTR) in 2002–15 was approximately $45.4 million,[3] only a portion of which is spent on WTO litigation.[4] When the USTR was threatened with budget cuts in 2013, its top litigator warned that his office "may no longer have the funding to initiate new legal disputes, which would result in reduced enforcement of trade agreements."[5] European trade politics experts suggest that the EU faces even more severe budgetary constraints on litigation (Greenwood 2000; Shaffer 2003a). WTO disputes also come with high political costs. Accusing a trade partner of a violation inevitably antagonizes it, putting diplomacy at risk. Japan's foreign affairs ministry, for instance, regularly prevents Japan's trade ministry from filing WTO complaints against China because diplomats fear that a WTO case would exacerbate Japan's ongoing foreign policy conflicts with China.[6] Small countries may be especially prone to such fears, especially if they rely on foreign aid (Elsig and Stucki 2011). Shaffer (2003a, 137) argues that such concerns also affect the United States and the EU, which must balance the benefits of trade litigation against the desire to maintain "friendlier relations" with their trading partners.

Despite these costs, WTO members frequently file disputes against their trade partners: WTO members have filed over 520 formal disputes since 1995. They do so primarily for economic reasons. By liberalizing global markets, states provide their exporters with access to foreign markets and consumers with access to a greater diversity of goods at more competitive prices. Recent studies disagree on the magnitude of the economic benefits of WTO litigation. Bechtel and Sattler (2015) find that a complainant's exports increase by over US$7 billion after panel rulings, while Chaudoin, Kucik, and Pelc (2016) and Bown and Reynolds (2015) find more ambiguous average effects. Yet economic benefits do not explain all WTO litigation. Filing decisions are shaped by a potential complainant's legal capacity and its retaliatory power vis-à-vis the defendant country (Bown 2005; Busch, Reinhardt, and Shaffer 2009; Davis and Bermeo 2009). Both factors reinforce existing power asymmetries in the international system, privileging the in-

---

3. Data from 2002–7 were provided by Fred Ames, the Assistant US Trade Representative for Administration, and data from 2008–15 came from the annual Congressional Budget Report.

4. The USTR does not track expenditures based on its activities, but average annual spending by the USTR's office in Geneva (where WTO litigation actually occurs) was only $6.4 million during 2008–15.

5. "Reif: Sequestration Could Hinder Litigation, Negotiating Efforts at USTR," *Inside US Trade*, March 1, 2013.

6. Author interviews with officials from Japan's Ministry of Economy, Trade and Industry on November 19, 2013.

terests of highly developed states at the expense of developing states (Bown and Hoekman 2005; Sattler and Bernauer 2011).

Scholars have also emphasized the importance of domestic politics in shaping decisions about whether to file WTO disputes (Chaudoin 2014; Davis 2003, 2012). Domestic groups that benefit from open markets have incentive to lobby their governments to expand and enforce international trade rules (Johns and Rosendorff 2009). Not all firms lobby equally; Davis and Shirato (2007) find that "low-velocity" Japanese industries with longer time horizons are more likely to press for enforcement. Scholars have also looked to domestic groups in the defendant country to account for the timing of disputes: Chaudoin (2014) argues that countries are more likely to challenge US trade measures during US election years with low unemployment. Our argument, by contrast, cuts across domestic political institutions, focusing on how a fundamental attribute of trade measures themselves—namely, their diffuseness—affects the odds that they will be challenged.

Even though litigants must bear the full cost of enforcement at the WTO, the distribution of the benefits of enforcement varies across potential cases. We argue that the diffuseness of a trade policy affects the degree to which litigation is a public good. Regardless of their total economic effect, some trade policies, like the 1998 US restrictions on Canadian trucks, affect only one country, meaning that the policy's impact is highly concentrated. For such concentrated policies, dispute settlement is largely a private good because a complainant internalizes most of the benefits of challenging a possible violation. Other trade policies, like US corn subsidies, affect a great many countries, meaning that these policies have a highly diffuse impact. For such diffuse policies, dispute settlement resembles a public good, generating "an incentive to free-ride on the litigation of others" (Bown 2005). Public goods generate collective action problems: when individual effort is needed to create a good that benefits all, each individual has an incentive to free ride on the effort of others (Olson 1965). As more actors benefit, each individual is less likely to provide it. A government's desire to free ride on other states affects whether domestic political groups can successfully pressure their government to enforce international trade rules. In our model, domestic political groups can more easily push governments into filing cases when the benefits of enforcement are highly concentrated. When a case generates a highly diffuse benefit, however, governments are less likely to be swayed into launching a costly dispute. Of course, no WTO dispute is ever a purely public good. Challenging a highly concentrated policy can yield some public benefits. Other countries may benefit indirectly by a dispute's spillover effects on jurisprudence (Pelc 2014). Litigation can also clarify the meaning of specific legal provisions, leading to less uncertainty in the future about the meaning of WTO law. Finally, all countries benefit in the long run if the WTO's dispute settlement system deters future violations (Johns 2012).

At the other extreme, challenging a highly diffuse policy can yield some private benefits. Under WTO rules, all members must provide most-favored nation (MFN) treatment to all other members, meaning that any concessions obtained through dispute settlement must be extended to all members.[7] So when a complainant successfully challenges an import restriction on apples, for example, all countries that export apples to the defendant stand to benefit from the removal of the restriction. Yet politics sometimes intervenes, leading litigants to craft discriminatory settlements that redirect at least some of the benefits of enforcement to those countries that participated in litigation (Bagwell and Staiger 2004; Davis 2003; Johns and Pelc 2014, 2016).[8] Nevertheless, the public benefits of litigation are thought to be large enough that trade scholars routinely refer to WTO dispute settlement as a public good (Bechtel and Sattler 2015; Bown 2005). In fact, the public benefits of WTO litigation are usually presented as a key positive aspect of the regime: litigation by a few (mostly powerful) countries benefits everyone, because it lifts protectionist barriers that may affect the membership as a whole. Yet we show that this feature of the trade regime also has overlooked consequences that are generated by underlying collective action problem.

Our theoretical model provides the logic for two main claims. First, more diffuse policies take longer to be challenged and risk not being challenged at all. Enforcement delays are highly costly to affected countries. Each day of delay is a day during which a distortionary policy remains in place, with the economic consequences it entails. Most often, the WTO litigation process does not yield the amount of alleged harm produced by the disputed policy. Yet some WTO cases go through an arbitration process in which damages are calculated.[9] To offer a rough sense of scale, the estimated average annual harm caused by each disputed policy across all such cases was US$622 million, or about US$1.7 million a day.[10]

---

7. The extension of the MFN principle to any concessions obtained through dispute settlement is found in Article 3.5 of the DSU: "All solutions to matters formally raised under the consultation and dispute settlement provisions . . . shall be consistent with those agreements and shall not nullify or impair benefits accruing to any Member."

8. In our theoretical model below, we account for such situations by assuming that litigants can sometimes receive private benefits from litigation.

9. These damages are assessed during Article 22.6 proceedings.

10. Horn and Mavroidis (2011) data. This number is not representative of the average WTO challenge, since it is measured only for those disputes where compliance following a ruling was not forthcoming. The takeaway is simply that delayed enforcement comes at a significant cost.

The aim of the dispute settlement understanding is to minimize such distortionary effects.

Second, free riding affects which specific violations are challenged. When a country chooses whether to challenge a trade policy at the WTO, it must perform a cost-benefit analysis. In the case of highly concentrated policies, the choice is simple: a country should file if the direct benefit of litigation to the country is higher than the cost. But if others also gain from enforcement, the calculation becomes more complex, since the country will hope to free ride on the filing by another government. Holding the total trade stake constant, diffuse policies thus face a higher barrier to enforcement: a potential case must have a higher chance of success, all else equal, for an affected country to file it. The result is a testable implication that capitalizes on the known selection process: conditional on being filed, cases that are more diffuse should be more likely to succeed in litigation, all else equal. If the selection of disputes is not occurring on the basis we describe, then we should not expect to see any difference in legal success associated with diffuseness.

Of course, other factors might affect legal success at the WTO. For example, if a state wishes to challenge a highly concentrated policy, perhaps it will invest more resources in litigation, seeking assistance from private law firms and working more closely with affected industries to support its legal claims (Brutger 2014; Shaffer 2003b).[11] All else equal, this logic would suggest that more concentrated cases are more likely to succeed at the WTO. Our argument and evidence suggest that the opposite is true. Alternatively, perhaps more diffuse cases are more likely to succeed because more countries are willing to contribute to the collective effort of mounting a case. Similarly, perhaps WTO panelists will be more likely to rule against the defendant if the defendant's alleged violation harms more countries (Johns and Pelc 2014). Under both of these scenarios, we would expect that more diffuse cases are more successful cases, which in turn would cause affected states to file them more quickly. These alternative explanations would support our second main empirical finding—more diffuse cases are more likely to succeed in litigation than less diffuse cases. But they would contradict our first main empirical finding—more diffuse policies face greater enforcement delays.

Both of our theoretical claims are based on observable implications of an unobservable process: selection. An ideal research design would identify all possible trade violations and then observe which policies are challenged at the WTO. Some scholars have attempted to approximate this research design by focusing on the behavior of specific countries and describing a prior stage in the selection of disputes. Davis and Shirato (2007) thus examine which disputes are eventually filed among the potential violations identified in the annual report of Japan's Ministry of Economy, Trade, and Industry. Similarly, Chaudoin (2014) uses US antidumping and countervailing petitions as a pool of potential violations and then examines which are more likely to be challenged and at what point in time. Both of these studies have yielded valuable insights about the domestic politics of trade disputes, yet by their own admission, both push the selection process down one level, rather than accounting for it fully. The question then becomes, what process has brought these measures to the attention of, for example, the Japanese trade ministry, and whether this process is itself prone to selection bias. Using a prior universe of cases also necessarily limits the researcher to examining individual countries, with the risk that the findings may not be generalizable to the behavior of other WTO members. In view of these methodological pros and cons, here we propose a different approach. Rather than trying to systematically control for selection effects—an inherently infeasible task given the size of the WTO and the scale and diversity of domestic trade policies—we design empirical tests that rely upon selection effects. We can only observe cases that are actually filed, so we test our theoretical argument using observable patterns of behavior that should be present in these observable cases if selection is occurring.

Our argument suggests that the suboptimal provision of enforcement might be solved if states could effectively coordinate their efforts (Johns 2012, 2015). States can, after all, communicate among themselves about their intentions to litigate. However, any attempt at sharing the cost of enforcement is likely to suffer from the usual problems of international coordination: countries will prefer that others invest first, and the value of enforcement may vary across cases and time, complicating attempts at "enforcement-trading."

Consider the WTO's two most powerful members: the United States and the EC (European Union). In his detailed comparative study of EC and US trade disputes, Shaffer notes that the EC and US "rarely collaborate in WTO litigation" for many reasons, including antagonism from EC-US trade disputes and differences in legal strategies, organizational culture, and domestic political institutions (Shaffer 2003a, 127). More importantly, Shaffer argues that the EC and United States recognize their strategic incentives "to 'free-ride' on the other's aggressive actions. The more passive party thereby benefits from enhanced market access while retaining friendlier relations with the foreign country for other purposes" (137). Similarly, Steinberg (1999) discusses the challenges of EC-US cooperation on Asian trade policy. He writes: "The European

---

11. We thank an anonymous referee for suggesting this possibility.

Union and United States face many trade problems with third countries that if resolved on an MFN basis would have the qualities of a public good. When one transatlantic power successfully acts alone to resolve such a problem, . . . the other power may simply free ride on the result" (Steinberg 1999, 217).

There have been a few attempts at cooperation in WTO litigation. In *Japan—Alcoholic Beverages*, the EC and United States agreed to jointly challenge Japanese liquor laws, but cooperation quickly collapsed when the EC and the United States disagreed about legal arguments (Shaffer 2003a, 134). Multinational firms have helped states to coordinate successfully a few times. In the 1990s, alcohol lobbying groups in Canada, the EU, and the United States coordinated litigation against Korean alcohol tariffs. Similarly, Nike pressured the EU and United States to challenge Argentinian measures on footwear and apparel (Shaffer 2003a, 140–41). While these anecdotes suggest that cross-national coordination is sometimes possible, descriptive statistics show no systematic burden sharing in enforcement. Fewer than 10% of WTO disputes involve multiple complainants.[12] These cases usually challenge policies that are highly politically salient, suggesting that they may be less affected by the economic incentives that drive our argument. Additionally, these complainants tend to challenge different aspects of trade policies: the precise incentives of co-complainants often differ.

While sustained coordination over enforcement is unlikely, any such dynamics should bias our empirical tests against the effect we identify. That is, if countries can successfully coordinate—through either alternation or cost sharing—then we should expect that more diffuse policies would be more likely to be challenged. Such cases should thus be filed more quickly and have lower odds of legal success than cases with a less diffuse effect. Both these implications directly contradict our empirical findings. Next, we develop the theory to derive these empirical expectations.

THEORY

We present an infinite-horizon game with discrete time ($t = 1, 2, \ldots$).[13] The game begins when n countries are harmed by a new policy that disrupts their trade. We let $\tau_i > 0$ denote country i's trade stake—this represents the magnitude of country i's harm from the new policy. Since we care about decisions to enforce WTO rules, we assume the new policy

is exogenous and focus on the behavior of affected countries (countries with $\tau_i > 0$).[14]

We assume that each affected country faces some domestic pressure to challenge the new policy, but the size of this pressure changes exogenously over time. Conditional on reaching period t, each country i privately learns its type, α_{it}, which represents the domestic pressure on i to challenge the policy in period t.[15] Then all countries must simultaneously decide whether to file a WTO dispute. If country i does not file, it receives the payoff $-\alpha_{it}\tau_i$ for period t, which can be interpreted as player i's political or economic cost from failing to enforce in period t. Parameter α_{it} therefore represents the unit cost of failing to enforce, and the magnitude of the overall cost depends on a player's individual benefit from enforcement, τ_i. So diffuseness can affect the level of stochastic pressure if it influences the individual trade stake of a player. If no country files, the game progresses to period $t + 1$. If at least one country files, then the dispute goes to the WTO and our model ends.

Because we care about filing decisions, we model WTO dispute settlement in reduced form.[16] When the dispute goes to the WTO, all affected countries benefit from having the case resolved. While the complainant might lose a panel ruling, it might alternatively win a ruling or negotiate a settlement in which the policy is partly or completely removed. As discussed above, the most-favored nation principle ensures that all affected countries—including those countries that did not file—benefit from such outcomes. We let parameter $r > 0$ represent the case quality, and payoff $r\tau_i$ represent country i's expected per period payoff from WTO dispute settlement. This parameter includes expectations about the likelihood and consequences of successful litigation. Higher quality cases are by definition more likely to yield procomplainant rulings.

We allow the complainant to receive additional private benefits from dispute settlement. These private benefits might come from discriminatory settlements or any other indirect benefits of litigation that only the complainant receives. We let parameter $b > 0$ represent the expected private benefits, and payoff $b\tau_i$ represent country i's expected per period private benefit if it files the case. Finally, we let parameter $c > 0$ represent the one-period litigation cost and assume that countries have the discount factor $\delta \in (0, 1)$. The appendix, available online, contains each country's expected utility for the possible infinite streams of these per period payoffs.

---

12. Descriptives from the data used in the analysis.

13. Here we present the basic intuitions about our causal mechanism and observable behavior. Technical readers can consult the appendix for a more mathematical presentation.

14. We discuss endogenous trade policy—initial decisions about whether and how to violate—in the conclusion.

15. Assumptions about the distribution of α_{it} are in the appendix.

16. The conclusion discusses a model extension with a more detailed dispute settlement system.

We initially solve the model for all possible distributions of trade stakes, generating propositions 1–4. This approach allows us to consider how changes in positive externalities affect enforcement decisions, an issue that we return to below. However, to identify the impact of diffuseness, we must make an assumption about the distribution of trade stakes for propositions 5–8. Namely, we hold the total impact of the trade policy on all countries, $\tau = \sum_i \tau_i$, constant, and then assume that each country's trade stake is an equal share of the total trade stake, $\tau_i = \tau/n$. When there are few affected countries (small n), the overall impact of the trade policy is concentrated. However, as the number of affected countries increases, the total impact of the trade policy is spread across more countries, making it more diffuse. This approach allows us to hold the total impact of the trade policy fixed when taking comparative statics. Diffuseness reduces each individual's incentive to enforce but also increases the number of players with some incentive to enforce. We are therefore isolating the impact of diffuseness, independent of the overall aggregate benefit of enforcement. We first identify a property of the weak perfect Bayesian equilibrium for our game.[17]

Proposition 1. When countries are relatively impatient (δ is small), each country adopts a cut point strategy: conditional on reaching t, high types file the case and low types do not file.

Given our model's structure, we must constrain the discount factor to identify a reasonable equilibrium. To understand why, suppose that the countries are extremely patient (δ is large). Then an infinite stream of even small expected private benefits will outweigh the one-period litigation cost, and all countries will immediately file the case. This behavior is substantively implausible because countries rarely file WTO disputes the moment new policies appear. More plausible behavior occurs when countries are relatively impatient (δ is small) because the one-period litigation cost deters some types from filing a case.

When a country has a small trade stake, both the expected benefit of filing the case and the domestic cost of not filing are small, so a country will not want to pay the litigation cost. Not surprisingly, a country is more likely to challenge a trade policy that harms it more.

Proposition 2. When its own trade stake (τ_i) increases, country i is more likely to file in any given period.

However, a country's incentives are different when the trade policy causes more harm for another country. When the trade stake of another country j increases, any enforcement action by country i generates larger positive externalities. The aggregate economic benefit of enforcement increases. Country i does not directly care about the trade stake of another country j. However, country j's trade stake indirectly affects country i by changing i's beliefs about how j will behave. Countries want the long-term benefit of going to the WTO but do not want to pay the short-term litigation cost. So country i is less likely to file when another country j is more likely to file, which occurs when j's trade stake increases. Greater positive externalities can therefore reduce enforcement at the individual level if they cause a player to believe that someone else is more likely to enforce.

Proposition 3. When another country's trade stake (τ_j) increases, country i is less likely to file in any given period.

Case quality, r, also affects a country's behavior. Since all countries benefit when the case is filed, the expected utility from both filing and not filing increases when case quality increases. However, the expected utility functions change at different rates. When a country files, it knows that its payoff is increasing as r increases. However, when a country does not file, an increase in r only benefits the country if someone else files the case. So increasing a case's quality makes filing the case more attractive relative to not filing.

Proposition 4. When the case quality increases, each country is more likely to file the case in any given period.

Propositions 2 and 3 suggest that the distribution of harm across affected countries matters. However, neither result isolates the effect of diffusiveness because by increasing the trade stake of one country in the results above, we are also increasing the total impact of the trade policy on all countries, $\tau = \sum_i \tau_i$. We now invoke our additional assumption about the distribution of trade stakes, which is described above. That is, we assume that $\tau_i = \tau/n$. Increasing the number of affected countries therefore decreases each individual player's benefit from enforcement because the total trade impact is spread across more players. When a trade policy's impact is spread across more countries, each country's individual trade stake decreases, exacerbating the collective action problem.

---

17. This solution concept requires that strategies are sequentially rational and beliefs are consistent with Bayes's Rule where possible. Since types are independent across time and players choose actions simultaneously, we do not need to specify off-the-equilibrium-path beliefs.

Each country is more tempted to free ride and therefore is less likely to file.

Proposition 5. When the number of affected countries increases, each country is less likely to file in any given period.

However, this individual-level effect does not necessarily extend to the collective outcome—whether someone files a case. At the individual level, diffuseness makes each country less likely to file because the total benefit of enforcement is distributed across more individuals. Yet diffuseness also increases the number of countries that want to file the case. Which effect is dominant—the individual versus the collective—depends on the model parameters. Suppose we increase the number of affected countries from n to $n + 1$. This spreads the impact of the trade policy across more countries, decreasing the likelihood that one of the original n countries will file. If the litigation cost is relatively small, the new country is likely to file the case, offsetting the decrease in the likelihood that one of the original n countries will file. However, as the litigation cost grows, the new country is less likely to file, and the negative impact of diffuseness on the original n countries outweighs the effect of increasing the number of affected countries.[18]

Proposition 6. When the litigation cost is large and the number of affected countries increases, the overall probability that the case is filed by at least one country decreases.

Under an ideal research design, we could identify all possible trade violations, observe which policies are challenged at the WTO, and examine whether diffuse policies are less likely to be challenged than concentrated policies. However, the available data have selection effects: we can only observe cases that are actually filed. Nevertheless, our model generates two major empirical implications—concerning enforcement delay and legal outcomes—that can be tested using observed legal challenges. Rather than being constrained by selection effects in our empirical analysis, we leverage selection effects in the model to generate hypotheses about observable disputes.

First, the results above concern the likelihood that a dispute is filed in a given period. But because we have an infinite-horizon game, we can make meaningful and rigorous inferences about duration—how long countries will wait to file a

case. We refer to this as "enforcement delay." Because violations with more diffuse effects are less likely to be challenged in a given period, diffuseness increases enforcement delay in observed disputes.

Proposition 7. In observable WTO disputes, cases that challenge more diffuse policies will, on average, have more enforcement delay, ceteris paribus.

This finding yields insight into a collective outcome, namely, how long it takes someone to file a dispute. But it also allows us to indirectly test our arguments regarding individual-level behavior, even in the presence of selection effects. As described below, statistical methods for estimating the duration of an outcome rely upon the use of hazard ratios. In our analysis, these hazard ratios indicate the likelihood that the case is filed in a given unit of time. If we construct our analysis at the dispute-country level, we can gain leverage over decision making by individual states. This allows us to assess the validity of propositions 2, 3, and 5. All of these results should hold when we restrict attention to those cases that are eventually filed and, hence, enter our sample.

Second, we showed that diffuseness and case quality have opposing effects on filing decisions: diffuseness reduces the likelihood that a case is filed, and legal merit increases this likelihood. When deciding whether to file a case, every country must balance the expected benefit from litigation against its expected cost. Suppose that a given country is indifferent between filing and not filing the dispute. If the number of affected countries increases, then the expected benefit of filing decreases. If we wish to offset this effect to ensure that the country remains indifferent, we must increase the quality of the case. So there is a clear selection effect: conditional on being filed, a case that challenges a diffuse policy should be of higher quality, on average, than a case that challenges a concentrated policy. Accordingly, it should be more likely to yield a procomplainant ruling.

Proposition 8. In observable WTO disputes, cases that challenge diffuse policies will, on average, be more likely to end in a procomplainant ruling than cases that challenge concentrated policies, ceteris paribus.

This argument is illustrated in figure 1. In the top half of figure 1, the x-axis represents case quality (r) and the y-axis represents a country's expected benefit from filing a dispute. Proposition 4 states that as case quality increases, a country's expected benefit from filing increases, as shown by the upward-sloping lines in figure 1. When the expected benefit is positive (above the dashed line), a country will

---

18. In the appendix, we show that this logic holds when $\alpha \sim U[0, A]$ and c is large. We also derive the necessary and sufficient condition on the distribution function for propositions 6–8.
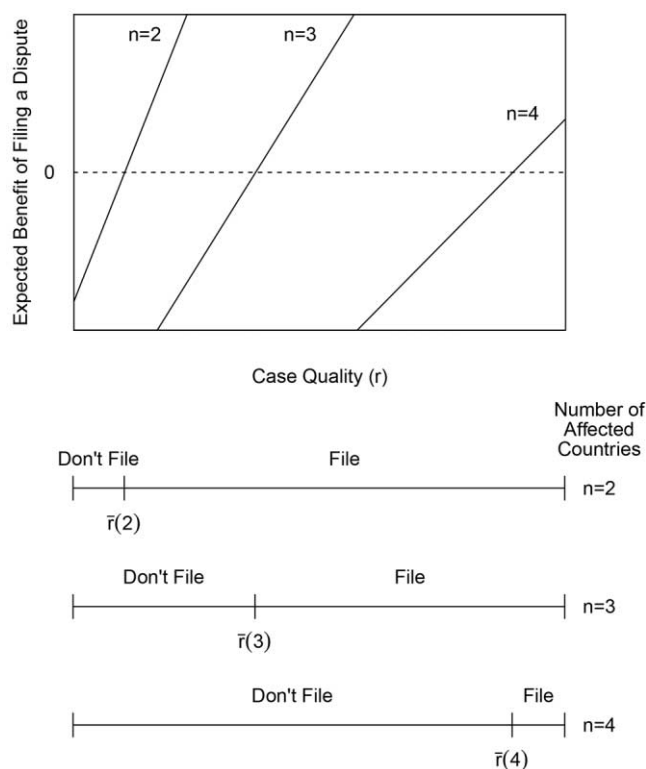
Figure 1. Diffuseness increases the quality of observed cases. Figure created from simulations of equilibrium behavior in R.

file; when the expected benefit is negative (below the dashed line), a country will not file. Proposition 5 states that increasing the number of affected countries makes the trade policy more diffuse, reducing an individual country's incentive to file. So the expected benefit of filing when there are four affected countries ($n = 4$) is lower than the expected benefit when there are three countries ($n = 3$), which in turn is lower than the expected benefit when there are only two countries ($n = 2$). The bottom half of figure 1 contains line graphs that show the minimum case quality needed for a country to want to file the case, $\bar{r}(n)$. When there are only two affected countries ($n = 2$), a player will file anytime that the case quality is at least as large as $\bar{r}(2)$. This is the critical value of r at which the expected utility from filing is equal to zero. Similarly, when there are three countries ($n = 3$), the critical value of case quality is $\bar{r}(3)$. Because the expected utility from filing is lower if there are three players than if there are two, the case must be better for a state to be willing to file ($\bar{r}(2) < \bar{r}(3)$). Finally, $\bar{r}(4)$ is the critical level of case quality when there are four affected countries ($n = 4$).[19]

Because both propositions 7 and 8 are explicitly driven by selection effects, they can be tested on observable cases. We can thus leverage the selection effects that are inherent in the data-generating process rather than being constrained by selection. Namely, we can test proposition 7 by examining how long trade policies were in effect before being challenged at the WTO. Similarly, we can test proposition 8 by examining the outcomes from actual WTO disputes. If the perceptions of countries are correct—that is, if higher quality cases are more likely to generate procomplainant rulings—then cases that challenge diffuse policies should be more likely to generate legal victories for the complainant than cases that challenge concentrated policies.

How robust are our findings?[20] We argue that the diffuseness of a policy is analytically distinct from its total aggregate effect, and we isolate the impact of diffuseness by holding the total aggregate effect of a policy constant and then spreading this effect over more players. Alternatively, we could hold each individual's trade stake constant and increase the number of affected players. This would confound the impact of diffuseness because it would increase the policy's aggregate effect. In this scenario, individuals have incentive to free ride, but the likelihood of enforcement increases because there are more possible enforcers and the policy has a larger effect. So when we test propositions 7–8 empirically, we must control for a policy's total trade stake. We could also assume that the unit public reward is a function of diffuseness.[21] States may benefit more from enforcing diffuse violations if they induce spillover effects, or generate more legal capacity (Shaffer 2003a). Similarly, WTO panelists may give more favorable rulings when a policy affects more states.[22] Under these scenarios, diffuseness would increase the unit value of the public reward. We show in the appendix that this analytical change weakens, but does not invalidate, our argument.[23] Our results also hold if we allow individual variation in litigation costs, or if the litigation cost decreases somewhat when the

---

19. As we show in the appendix, similar logic suggests that in observable WTO disputes, cases that challenge diffuse policies will, on average, involve higher total trade stakes. The data described below support this

claim, as there is a positive and significant correlation between our diffuseness measures and the total trade stake of a dispute. We thank an anonymous reviewer for noticing this fact.

20. All of the model extensions discussed here are included in the appendix.

21. We thank an anonymous referee for the many suggestions considered here.

22. For competing arguments on bias in WTO rulings, see Busch and Pelc (2010) and Johns and Pelc (2014).

23. The impact of diffuseness will depend on the values of exogenous parameters, but our results hold when states are relatively impatient, private benefits are relatively small, or the litigation cost is relatively large. We believe that these conditions hold in the context of the WTO.

policy affects more states.[24] Finally, if WTO panels are biased in favor of cases that challenge more diffuse policies, then more diffuse policies will generate more procomplainant rulings, consistent with proposition 8; but diffuseness will increase the expected utility of filing, thereby reducing enforcement delay, contradicting proposition 7.

EMPIRICS

Testing our two hypotheses requires two distinct data sets. Our first data set is built at the dispute-country level and includes one observation for each potential challenger of a given trade measure for every WTO dispute since 1995. Our second data set collapses these data to the dispute level and considers the proportion of claims ruled in favor of the complainant in each WTO dispute.

Free riding and enforcement delays

To test our first hypothesis, we need a measure of Enforcement Delay—the time it took for a complainant to file each WTO dispute. Bown and Reynolds (2015) includes the implementation date for the trade policy underlying every WTO dispute. We compare this implementation date to the date on which each WTO dispute was filed. This allows us to measure, in days, exactly how much time elapsed between the start of a policy and its challenge. We start the clock at the WTO's inception because of the change in countries' obligations between the GATT and the WTO period, and exclude countries that were not WTO members at the time of a dispute's initiation.[25]

We use three measures of diffuseness. For our first measure of diffuseness—Number of Countries Affected—we identify which products are affected by the trade policy being challenged, then take the log of the number of countries with more than US$500,000 in trade at stake in the year that the dispute began.[26] Our second measure of diffuseness—Disputed Trade Flows HHi—captures the distribution of trade. We construct a Herfindahl-Hirschman index (HHi) measure of trade flows into the defendant country.[27] The greater (smaller) this variable, the more concentrated (diffuse) trade

in the disputed product is. These two measures are highly negatively correlated, so our estimations feature only one of these two variables at a time. Both of these variables use bilateral trade flow data from the World Integrated Trade Service, which inevitably leaves out nonmerchandise disputes for which we cannot quantify the amount of trade. We code such cases as missing for the purpose of our first two diffuseness measures.

Our third measure of diffuseness—Global Policy—includes both merchandise and nonmerchandise disputes. This dichotomous variable comes from Bown and Reynolds (2015) and indicates whether a trade policy is "global," meaning that it affects the entire membership, or "partial," meaning that it affects only a subset of members.[28] A case is coded as a global policy if all foreign countries that exported the underlying good were affected. On average, partial policies are challenged after 779 days, while global policies are challenged after 1,034 days—a statistically significant difference.[29]

In cases such as our opening example of the US measure blocking Canadian trucks, a variable indicating concentration of trade flows in the disputed product (cattle and swine) lacks relevance. The legal nature of the policy is logically prior. Yet disputes like the Canadian trucking case are rare: few trade measures truly concern a single country, even among those coded as partial. Most partial disputes affect a significant subset of the membership, in which case economic diffuseness variables remain useful indicators. We thus test our two economic diffuseness variables both on a sample restricted to disputes that challenged global policies and on the entire caseload.

Our first control variable, Own Trade Stake, is the log of exports of the disputed product from the country under observation into the defendant market. Our second control variable, ROW Trade Stake, is the log of the rest of the world's (ROW) exports of the disputed product into the defendant's market—that is, everyone but the country under observation. We also control for legal capacity. The more countries participate in the dispute settlement process, the more likely they appear to challenge policies that affect them (Davis and Bermeo 2009), so we code Country Legal Experience as a running count of the number of prior cases a country has filed. In our main estimations, we include a measure of logged GDP, GDP per capita, and trade dependence for both countries in each case. These are intended to capture any effect of retaliatory power and can also be interpreted as an alternative measure of legal capacity. Finally, we control for the year in

---

24. The precise condition on changes in cost is specified in the appendix. If this occurs in some cases, our empirical tests would be biased against our theory.

25. Our results are statistically and substantively stronger if we consider the GATT implementation dates instead, likely owing to greater variance. We thus show more conservative findings below.

26. This count varies from 1 to 95, and its graph is in the appendix. We verify that the core findings are not sensitive to this choice in the appendix.

27. The HHi measure is $HHi = \sum_{i=1}^{n} f_i^2$, where f_i is the trade share from country i to the defendant in the product at issue, and n is the number of countries with nonzero trade to the defendant in the product at issue.

28. About 48% of the cases in our sample involve global protectionist policies, and 52% involve partial policies.

29. In the appendix, we plot the distributions of each type of case.

which a dispute was initiated, to account for potential trends across time. We include fixed effects for the 19 legal issues in our data to ensure that our results are not being driven by the inherent differences between our legal measures, which could affect the delay in filing.[30]

We estimate a Cox proportional hazards model, shown in table 1. Recall that the data are at the dispute-country level of observation: they include information not only about the country that eventually filed but also about all the countries that did not. We thus right-censor any countries that did not challenge the violation to account for how, had the violation not been challenged by the eventual complainant, these countries may still have done so.[31] Also recall that our assessment of economic diffuseness is limited to those disputes where we can observe trade flows in the disputed product. This leaves out disputes that concern nonmerchandise issues. Yet these still vary along our legal diffuseness variable, Global Policy. We thus begin our analysis with a parsimonious Cox proportional hazards model that considers only the relationship between Global Policy and the likelihood the measure is challenged in any given period, controlling for Country Legal Experience and the Initiation Year, which do not require trade flows data. The estimation, shown in model 1 of table 1, thus exploits our maximal sample.[32] Model 2 adds our set of trade and market control variables, which restricts our analysis to nonmerchandise disputes. Models 3 and 4 of table 1 estimate the effect of our two economic measures of diffuseness, Number of Countries Affected and Disputed Trade Flows HHi, in succession. We convert hazard ratios into coefficients; a negative coefficient represents a decrease in the hazard function, meaning that a challenge is less likely in any given period, resulting in a longer enforcement delay.[33]

What does table 1 tell us? The type of legal discrimination has a consistent effect: Global Policy is associated with a significant decreased rate of legal challenge. The effect is highly substantively significant throughout. As an example, in model 1, the rate of legal challenge decreases by 38% when the legal violation at issue is a global policy. Our two economic indicators of diffuseness follow expectations. The greater the number of countries with a stake in challenging the policy, the longer such challenges take on average. Controlling for the number of countries, and the trade at stake for both the country under observation and the rest of the world, the concentration of trade flows pertaining to the challenged policy also has the expected effect: the positive coefficient on Disputed Trade Flows HHi shows that the higher the HHi, meaning the more concentrated trade flows pertaining to the protectionist policy are across members, the shorter the expected delay before a policy is challenged. Specifically, an increase of the HHi by one standard deviation (0.23) increases the rate of filing by 22%. Both indicators thus support the same belief about the concentration of benefits: more diffuse policies generate a longer enforcement delay. We also graph the cumulative hazard associated with a concentrated effects policy versus a diffuse effects policy, holding all else equal, in figure 2.[34] As the figure makes clear, more diffuse policies face a considerably lower rate of challenge than more concentrated policies.

Our two trade stakes variables prove equally interesting. Recall that under proposition 2, we expect that increasing a country's Own Trade Stake makes the country more likely to file all else equal, reducing enforcement delay. As expected, table 1 shows that the more trade a given country has at stake, the greater the hazard rate and the shorter the enforcement delay. In contrast, proposition 3 suggests that increasing the trade stake of other countries will decrease a country's willingness to file, thereby increasing delay. As per table 1, ROW Trade Stake is consistently and significantly negatively related to the hazard rate (except in model 4, where it remains negative but falls short of significance), meaning that all else equal, violations where the rest of the world has more at stake result in a greater enforcement delay. This relationship also holds in a univariate estimation. This finding conveys one striking implication of the free rider problem: more serious violations—as measured by trade flows in the rest of the world, from the point of view of each country—have a lesser rate of challenge and thus generate longer enforcement delays, on average.[35] In sum, enforcement choices appear individually, if not socially, rational.

Prior legal experience increases the likelihood of a challenge, but few of the market size variables in table 1 appear

---

30. In the appendix, we also include shared frailty on the legal issue.

31. Note that this makes our data set different from most survival data, since a supermajority of our observations are censored. Only a handful of violations are challenged by more than one party, sometimes not simultaneously. Given the unusual structure of the dispute-country level data, however, we later rerun the analysis at the dispute level.

32. We generate observations for all WTO members for each nonmerchandise measure in the data set.

33. In the appendix, we graph the scaled residuals for each of our three diffuseness variables against time. In all three cases, there is little indication of any time trend, suggesting that the treatment hazards ratio is pretty constant and that the proportional hazards assumption holds.

34. In figure 2, we assume that concentrated (diffuse) policies are partial (global) policies with one standard deviation below (above) the average number of countries affected, and one standard deviation above (below) the average trade HHi.

35. If we construct a total trade at stake variable, it too is negatively related to the odds of filing.

Table 1. Diffuseness of Violations and the Rate of Legal Challenge

	(1)	(2)	(3)	(4)
Global Policy	−.48***	−.97***	−1.00***	−.76***
	(.13)	(.21)	(.21)	(.19)
Disputed Trade Flows HHi			1.08***	
			(.32)	
Number of Countries Affected (log)				−1.20***
				(.16)
Own Trade Stake (log)		.36***	.34***	.42***
		(.04)	(.04)	(.04)
ROW Trade Stake (log)		−.22***	−.19***	−.04
		(.04)	(.04)	(.05)
Country GDP/cap (log)		−.26***	−.26***	−.28***
		(.07)	(.07)	(.08)
Country GDP (log)		−.05	−.04	−.10
		(.08)	(.08)	(.07)
Country Trade Dependence		−.00*	−.00*	−.00*
		(.00)	(.00)	(.00)
Defendant GDP/cap (log)		−.04	−.08	−.12
		(.11)	(.12)	(.10)
Defendant GDP (log)		−.19***	−.16**	−.06
		(.06)	(.07)	(.07)
Defendant Trade Dependence		−.01*	−.00	−.00
		(.00)	(.00)	(.00)
Country Legal Experience	1.23***	.54***	.52***	.54***
	(.04)	(.10)	(.10)	(.10)
Initiation Year	−.24***	−.14***	−.14***	−.15***
	(.03)	(.02)	(.02)	(.02)
Legal issue fixed effects		Yes	Yes	Yes
N	29,488	16,145	16,145	16,145

Note. Cox proportional hazards estimates, errors clustered on common dispute.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

to have consistent effects. Larger economies appear to be challenged at a somewhat lower rate. And wealthier countries appear less likely to challenge a measure in any given period once we control for legal experience, but the variable shows a positive coefficient (meaning a higher rate of challenge) in a univariate estimation.

We seek to ensure that the relationship we have identified between diffuseness and enforcement delay is a robust one.[36] One concern might be that our estimations consider the entire membership, while we know that only a subset of countries actively file legal challenges. We have included the Country Legal Capacity to account for such variation in the ability to file, but we also go a step further by reestimating

our models, examining only those countries that have filed a WTO complaint of their own in the past. We also ensure that disputes with multiple complainants are not driving the results. We do this first by adding a control for the number of complainants in each dispute and then by restricting the sample to single-complainant disputes. We also add control variables for EU and US partners, in case these members are inherently more likely to file challenges. We also tweak one of our key explanatory variables by varying the amount of trade required to count a country as being "affected by the measure."[37] We also rerun all our estimations using enforcement delays that stretch back into the GATT period. We also modify the "enforcement clock" for agricultural disputes. Be-

36. The results described in this paragraph are available in the appendix.

37. We test thresholds of USD0 (i.e., any positive trade), US$100,000, our baseline threshold of US$500,000, and US$1 million.
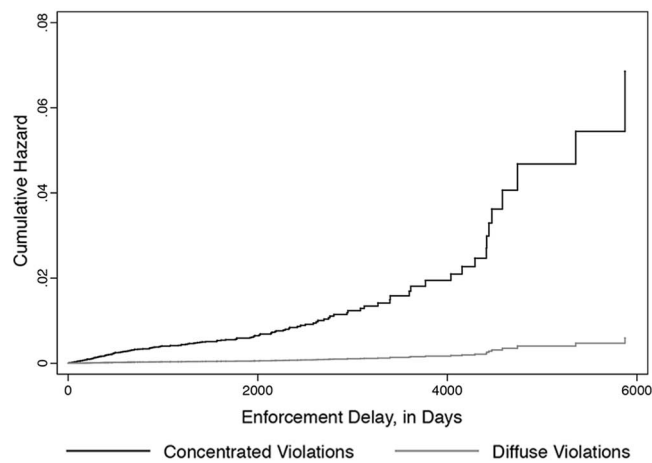
Figure 2. Cox proportional hazards regression

cause countries were given additional time to implement their new agricultural obligations under the WTO, we start the clock for agricultural disputes at January 31, 2000, the final deadline for all countries. This is an approximate adjustment, since agricultural disputes were also being fought in the first years of the WTO. Most drastically, we also test our argument by collapsing the data set to the dispute level, to provide a precise test of proposition 7, and to ensure that our findings are not driven by the unusual structure of the data, where a majority of observations are censored. Our theoretical expectations continue to hold after these modifications. Diffuse cases result in longer enforcement delays, and the more other countries care about a violation, the less likely a given country is to challenge it swiftly.

Does the diffuseness of benefits from litigation affect dispute outcomes?

Having found support for our beliefs about enforcement delay, we test the implications of this selection process on the likelihood of procomplainant rulings, as per proposition 8. It is well known that WTO disputes display a procomplainant bias: most rulings find some violation at play. Indeed, this fact is consistent with our theory: international trade rules are sufficiently clear that countries are able to gauge odds of legal success and choose which cases to challenge accordingly. Yet this procomplainant bias conceals quite a bit of variation, as complainants file a number of claims in a given dispute, allowing us to measure exactly how much of a dispute was ruled in favor of the complainant.

To construct our variable, we code the direction of every claim in every WTO dispute from 1995 to 2013.[38] This is a con-

siderable coding exercise: complainants have brought 4,484 such claims over the WTO's history. Of these, a minority are actually ruled on,[39] though some claims receive more than one finding. All told, panels have delivered 1,429 findings on 820 individual claims. We first collapse these findings at the claim level and then collapse claims at the dispute level to obtain the number of claims won by the complainant. We divide this number by the total number of claims filed to obtain the proportion of claims won by the complainant.

Most rulings are appealed, and the Appellate Body (AB) frequently overturns panel rulings. Unlike panels, which rely on ad hoc judges, the AB is a standing body, and its rulings are thus thought to have greater authority than panel rulings (Bhala 1999; Kucik and Pelc 2016). As a result, we are interested in rulings "net of appeal." The resulting dependent variable, Ruling Won Net of Appeal, considers the panel ruling, as modified (or not) by the AB. Keeping with existing work, we code this measure of legal success as a binary indicator, where procomplainant disputes are those where 90% or more of the claims were ruled procomplainant.[40]

We estimate the success of complainants using a Heckman selection model. The purpose is to account for the selection of cases that reach the ruling stage: about half of all disputes never make it to a ruling, and this risks biasing our results. We thus begin by estimating the odds of a ruling in a first-stage equation and use those estimates in our second-stage outcome equation. While we argue that a selection model is called for in this case, the results remain when we use a one-stage estimation that does not correct for selection.

Our explanatory variables of interest remain as described above in "Free Riding and Enforcement Delays." A simple descriptive statistic supports our expectations: global cases result in a procomplainant ruling significantly more than partial cases (0.48 vs. 0.32), and this difference is highly statistically significant. Of course, this relationship neither accounts for selection nor controls for confounding factors. To identify our model, we use the Number of Third Parties in the room. Many studies have shown that the presence of third parties decreases the odds of settlement and increases the odds of litigation, in part because litigants have an incentive to posture for the sake of third countries (Busch and Reinhardt 2006; Kucik and Pelc 2013). Yet the number of third parties in the room should not, by itself, have a direct effect on the direction of the ruling. Countries become third parties

---

38. A "claim" is an alleged violation of a given Article or subarticle of the WTO texts. These are taken directly from the complainants' request for consultations. There is thus no room for ambiguity in the coding.

39. For instance, all cases alleging national treatment discrimination make a claim under GATT III, yet the panel rarely rules on GATT III and, instead, rules on a specific agreement, such as the Agreement on Sanitary and Phytosanitary Measures.

40. In our usable sample, 44% of disputes fall into this category.

for a host of reasons, either because they care about the disputed product or because they care about the ruling's impact on jurisprudence. Third parties may join in support of both the complainant and the defendant (Busch and Pelc 2010). Countries also join simply to learn: following its accession, China joined nearly all disputes as a third party. Finally, recall that the expectation of legal success in the WTO does not increase with the amount of harm a measure may cause. In sum, we have little reason to believe that the number of third parties would have a direct effect on the direction of the ruling, and indeed, the variable is insignificant if included in the second-stage equation.

We control for the amount of trade the complainant has at stake and market size indicators—Complainant GDP (logged) and Defendant GDP (logged)—which may exert an impact at both stages of the estimation. We also control for Complainant Legal Experience and the Initiation Year. Anecdotal evidence suggests that in very large stakes, defendants cannot allow themselves to concede for domestic reasons without the "political cover" of an unfavorable ruling (Allee and Huth 2006). We thus expect Own Trade Stake to be positively related to the odds of a ruling. We cluster robust standard errors on the common dispute.

Our results, shown in table 2, provide strong support for proposition 8. Disputes over global policies fare better on average, and the greater the number of countries with exports at stake, the more successful the case, on average. Conversely, the more concentrated trade across those countries, the worse the prospects of the case, though this negative effect falls short of significance in the final model 4, where all three indicators are included simultaneously. Yet even in that model, the three concentration variables, taken together, remain highly jointly significant. These effects are substantively important. Looking at our first model, disputes over global policies are 77% more likely to result in a procomplainant ruling than disputes over partial policies,[41] and policies that rank as diffuse on all three indicators are more than twice as likely to result in a procomplainant ruling than the average dispute and nearly four times as likely as a concentrated dispute.[42] In short, disputes where the benefits of enforcement are more concentrated appear to be worse cases, on average.

The amount of trade at stake for the complainant increases the odds of litigation, in accordance with intuition, but has no consistent effect on the success rate of cases. The litigants' market sizes and the complainant's legal experience also have little consistent effect. Most importantly, the variable that iden-

tifies the model, Number of Third Parties, bears the expected strong positive relationship with the odds of litigation. Accounting for the selection of disputes into litigation, it appears that disputes over more diffuse policies are more successful cases. We obtain equivalent results in a model that omits selection altogether: a single-stage model yields qualitatively identical estimates.[43] But the selection model is common practice (Busch and Reinhardt 2006; Pelc 2017) and appears warranted.[44] We also ensure that the results are robust to modifications similar to those we made to our duration analysis: we add controls for the number of complainants in the dispute and then exclude all multiple complainants disputes altogether. In sum, various types of models all show that challenges of diffuse violations are simply better cases.

CONCLUSION

The enforcement of international agreements is often a public good. Yet in institutions that rely on decentralized enforcement, individuals must bear the private cost of enforcement, regardless of its aggregate benefits. In the WTO, legal challenges of highly diffuse policies approximate a public good because litigation benefits many countries. In contrast, a concentrated protectionist policy affects few countries, making enforcement a largely private good. We argue that as the diffuseness of a trade policy increases, so does the incentive for an affected country to free ride on enforcement by others, ceteris paribus. Put simply, more diffuse protectionism generates a more severe collective action problem that affects not only the likelihood of enforcement but also which kinds of violations are challenged.

The evidence supports our argument's two empirical implications. First, our theory suggests that diffuse policies should experience more enforcement delay. We measure diffuseness in three ways, through (i) the number of countries with trade at stake, (ii) the distribution of trade flows across those countries, and (iii) the legal nature of the violation at issue. On all three of these indicators, we find that more diffuse policies are associated with a longer delay between a protectionist policy's implementation and its eventual challenge. Cases over "global" legal issues, that concern a greater potential number of WTO members, are challenged at a 39% slower rate, representing considerable enforcement delay. We also find evidence supporting our expectation that while a country's own stake in the dispute should increase its willingness to swiftly challenge a protectionist policy, when others' stake in the dispute rises, that country becomes less likely to challenge the

---

41. Specifically, the predicted odds of a procomplainant ruling go from 19.4% to 34.5%, with other variables at their sample means.

42. See n. 34.

43. The results described here are available in the appendix.

44. A Wald test suggests that the correlation between the errors in the two equations confirms that these are not independent.

Table 2. Concentration of Benefits and Legal Success

	(1)	(2)	(3)	(4)	(5)
Ruling won net of appeal (second-stage equation):					
Global Policy	1.07**			1.12**	1.09**
	(.50)			(.47)	(.48)
Disputed Trade Flows HHi		−1.03**		−1.09**	
		(.52)		(.50)	
Number of Countries Affected (log)			.32*		.35**
			(.18)		(.16)
Own Trade Stake (log)	−.07	−.05	−.07	−.09*	−.12**
	(.05)	(.04)	(.05)	(.05)	(.05)
Complainant GDP (log)	−.04	−.02	−.06	−.01	−.05
	(.10)	(.10)	(.09)	(.10)	(.09)
Defendant GDP (log)	.11	.05	.03	.07	.05
	(.08)	(.08)	(.08)	(.09)	(.08)
Complainant Legal Experience	.10	.16	.18	.12	.15
	(.16)	(.15)	(.15)	(.16)	(.15)
Initiation Year	−.00	−.03	−.02	−.01	−.01
	(.03)	(.03)	(.03)	(.03)	(.03)
Constant	4.27	44.60	30.43	21.83	10.82
	(62.06)	(58.89)	(54.55)	(69.57)	(56.59)
Dispute goes to ruling (first-stage equation):					
Number of Third Parties	.25***	.25***	.25***	.25***	.25***
	(.05)	(.04)	(.04)	(.04)	(.04)
Own Trade Stake (log)	.05**	.05**	.05**	.05**	.05**
	(.02)	(.02)	(.02)	(.02)	(.02)
Complainant GDP (log)	−.00	−.01	−.01	−.01	−.01
	(.07)	(.06)	(.06)	(.06)	(.06)
Defendant GDP (log)	.06	.07	.07	.06	.06
	(.05)	(.04)	(.04)	(.04)	(.04)
Complainant Legal Experience	.11	.11	.11	.11	.10
	(.12)	(.10)	(.10)	(.10)	(.10)
Constant	−3.63	−3.58**	−3.58**	−3.59**	−3.59**
	(2.25)	(1.74)	(1.72)	(1.75)	(1.73)
Legal issue fixed effects	Yes	Yes	Yes	Yes	Yes
N	310	310	310	310	310

Note. Heckman probit selection model with maximum likelihood (ML) estimates. First-stage estimates likelihood of a ruling. Second-stage estimates likelihood of a procomplainant ruling. Robust standard errors clustered on the common dispute.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

policy in any given period. Second, our theory suggests that diffuseness increases the minimum expectation of legal success necessary for a country to want to file a dispute. We provide statistical evidence that supports this implication: cases that challenge more diffuse policies are considerably more likely to generate a procomplainant ruling. In fact, challenges of economically diffuse policies are associated with four times greater odds of a procomplainant ruling than challenges of concentrated policies, all else equal.

Our argument is not unique to the WTO. Most international dispute settlement bodies—including human rights bodies, investment tribunals, and regional trade agreements—rely upon decentralized enforcement. One avenue for future research is to look for evidence of free riding in institutions besides the WTO. Consider the European Union, which features a hybrid enforcement system, in which individuals, firms, member states, and the EU itself all have the ability to challenge possible violations of EU law. The conventional wis-

dom among EU scholars is that EU law has been enforced and developed primarily through lawsuits filed by individuals and firms, rather than by states and EU bodies (Alter 2003; Kelemen 2011). Yet one area of EU law is enforced almost exclusively by the EU itself: environmental regulation. Roughly 7.5% of the lawsuits brought by the Commission against member states in 1954–2009 involved environmental policies, while only 1.8% of the lawsuits brought by individuals and firms involved the environment.[45] Kelemen's explanation for this pattern matches our own: "most environmental regulation concerns matters of diffuse public interest. . . . Private parties often lack the individual incentive to commerce legal action to secure enforcement" (Kelemen 2004, 49). Johns (2016) suggests that similar patterns are apparent in other EU issue areas.

One striking implication of our findings is that if our theoretical argument is correct, countries should want to spread the pain of treaty violations as much as possible. In the realm of international trade, countries have an incentive to use diffuse policies when possible to avoid legal challenges, rather than using trade policies with a concentrated impact, such as countervailing duties and antidumping duties. This incentive might help us understand changes in the nature of trade violations. The growing frequency and importance of WTO litigation since 1995 has coincided with an increase in the use of standards as a tool for trade protectionism (Kim 2012; Kono 2006). These policies—such as labeling requirements, health and safety standards, and environment regulation—have highly diffuse effects, affecting all trading partners alike. Many factors naturally go into the design of import relief, but our analysis suggests that one benefit of protectionist standards may be that these are less likely to be challenged at the WTO than policies with a more concentrated effect.

A second implication of our findings pertains to potential cases that we do not observe. The collective action problem means that some protectionist policies may never be challenged because they affect many countries. Additionally, we show that when a country challenges a concentrated policy, it often does so at the expense of legal success. This suggests that limited resources are being spent on the "wrong cases." All else equal, affected countries would be better off if the resources that are spent on relatively weak cases with a concentrated impact were instead spent on stronger cases with a more diffuse impact.

The sheer difficulty of coordinating on enforcement suggests that in those rare cases in which it occurs—like in the Nike lawsuit mentioned above—filing decisions may be driven by factors other than instrumental trade concerns. While this lies outside of the scope of our argument, we suspect that coordinated filings can occur when WTO members perceive that there has been a gross violation of WTO rules, as occurred following the US imposition of steel safeguards in 2002 (Davis 2012; Pelc 2009). In such cases, WTO members may be willing to contribute to a collective enforcement effort (rather than free riding) in order to send a costly signal to domestic and international audiences that a large group of states are strongly opposed to the defendant's actions (Johns and Pelc 2014). Such cases remain very rare. Nevertheless, the symbolic value of litigation may sometimes help states to coordinate on enforcement.

Another factor that might help states to overcome the collective action problem is coordination by transnational actors, like multinational corporations. In international trade, most public-private partnerships occur within the state: firms and industry groups pressure government agencies to enforce trade rules and provide legal and financial assistance to do so (Bown 2009; Brutger 2014; Shaffer 2003a). However, the growth of multinational corporations suggests that private transnational actors may also be effective at coordinating enforcement by multiple governments, as in the aforementioned *Korea—Alcoholic Beverages* and Argentinean apparel disputes.

These cases also demonstrate how private transnational efforts at coordination are themselves subject to free-riding problems. In the case of Argentinean barriers on footwear, Nike, rather than a coalition of clothing or shoe manufacturers, ultimately bore the cost of enforcement coordination. When Argentina restricted footwear imports in 1997 through a range of measures, Nike was the global behemoth of shoe exporters.[46] When it bore the cost of coordinating WTO litigation, other shoe manufacturers, like Reebok and Adidas, were able to free ride on the benefits provided by Nike. Trade associations may be able to help individual firms to coordinate, but these organizations are themselves focused on lobbying for domestic regulation and hence organized at the national level. As Shaffer (2003a) documents, EC and US exporters created a transatlantic business dialogue in the late 1990s to try to better coordinate cross-national litigation at the WTO. These efforts ultimately proved unsuccessful.

In the case of both countries and firms, collective action affects not only delays in enforcement but also which legal violations are challenged and which are not. Overall, the temptation to free ride on enforcement means that litigation by a few cannot fully serve the interests of all.

---

45. Data are from the EUROPA database and are available from the authors.

46. In that year, Nike controlled over 35% of the global market share in athletic footwear. Its closest competitors were Reebok and Adidas, which controlled 14.5% and 10.3% of the global market, respectively (Locke 2002).

ACKNOWLEDGMENTS

REFERENCES

Allee, Todd L., and Paul K. Huth. 2006. "Legitimizing Dispute Settlement: International Legal Rulings as Domestic Political Cover." *American Political Science Review* 100 (2): 219–34.

Alter, Karen J. 2003. "Resolving or Exacerbating Disputes? The WTO's New Dispute Resolution System." *International Affairs* 79 (4): 783–800.

Bagwell, Kyle, and Robert W. Staiger. 2004. "Multilateral Trade Negotiations, Bilateral Opportunism and the Rules of GATT/WTO." *Journal of International Economics* 63 (1): 1–29.

Bechtel, Michael M., and Thomas Sattler. 2015. "What Is Litigation in the World Trade Organization Worth?" *International Organization* 69 (2): 375–403.

Bhala, Raj. 1999. "The Myth about Stare Decisis and International Trade Law (Part One of a Trilogy)." *American University International Law Review* 14 (4): 845–956.

Bown, Chad P. 2005. "Participation in WTO Dispute Settlement: Complainants, Interested Parties, and Free Riders." *World Bank Economic Review* 19 (2): 287–310.

Bown, Chad P. 2009. *Self-Enforcing Trade: Developing Countries and WTO Dispute Settlement.* Washington, DC: Brookings Institution.

Bown, Chad P., and Bernard M. Hoekman. 2005. "WTO Dispute Settlement and the Missing Developing Country Cases: Engaging the Private Sector." *Journal of International Economic Law* 8 (4): 861–90.

Bown, Chad P., and Kara M. Reynolds. 2015. "Trade Flows and Trade Disputes." *Review of International Organizations* 10 (2): 145–77.

Brutger, Ryan. 2014. "Screening for Success: The Effect of Firm Signaling on WTO Case Selection." Working paper, Princeton University.

Busch, Marc L., and Krzysztof J. Pelc. 2010. "The Politics of Judicial Economy at the World Trade Organization." *International Organization* 64 (Spring): 257–79.

Busch, Marc L., and Eric Reinhardt. 2006. "Three's a Crowd: Third Parties and WTO Dispute Settlement." *World Politics* 58 (April): 446–77.

Busch, Marc L., Eric Reinhardt, and Gregory Shaffer. 2009. "Does Legal Capacity Matter? Explaining Dispute Initiation and Antidumping Actions in the WTO." *World Trade Review* 8:559–77.

Chaudoin, Stephen. 2014. "Audience Features and the Strategic Timing of Trade Disputes." *International Organization* 68 (4): 877–911.

Chaudoin, Stephen, Jeffrey Kucik, and Krzysztof Pelc. 2016. "Do WTO Disputes Actually Increase Trade?" *International Studies Quarterly* 60 (2): 294–306.

Davis, Christina L. 2003. *Food Fights over Free Trade: How International Institutions and Issue Linkage Promote Agricultural Trade Liberalization.* Princeton, NJ: Princeton University Press.

Davis, Christina L. 2012. *Why Adjudicate? Enforcing Trade Rules.* Princeton, NJ: Princeton University Press.

Davis, Christina L., and Sarah Blodgett Bermeo. 2009. "Who Files? Developing Country Participation in GATT/WTO Adjudication." *Journal of Politics* 71 (3): 1033–49.

Davis, Christina L., and Yuki Shirato. 2007. "Firms, Governments, and WTO Adjudication: Japan's Selection of WTO Disputes." *World Politics* 59 (2): 274–313.

Elsig, Manfred, and Philipp Stucki. 2011. "Low-Income Developing Countries and WTO Litigation: Why Wake Up the Sleeping Dog?" *Review of International Political Economy* 19 (2): 292–316.

Greenwood, Justin. 2000. "Organized Business and the European Union." In Justin Greenwood and Henry Jacek, eds., *Organized Business and the New Global Order.* New York: St. Martin's, 77–98.

Horn, Henrik, and Petros Mavroidis. 2011. "The WTO Dispute Settlement Data Set 1995–2011." World Bank. http://go.worldbank.org/X5EZPHXJY0 (accessed February 22, 2017).

Johns, Leslie. 2012. "Courts as Coordinators: Endogenous Enforcement and Jurisdiction in Adjudication." *Journal of Conflict Resolution* 56 (2): 257–89.

Johns, Leslie. 2015. *Strengthening International Courts: The Hidden Costs of Legalization.* Ann Arbor: University of Michigan Press.

Johns, Leslie. 2016. "The Design of Enforcement: Collective Action and the Enforcement of International Law." Working paper, University of California, Los Angeles, Department of Political Science.

Johns, Leslie, and Krzysztof J. Pelc. 2014. "Who Gets to Be in the Room? Manipulating Participation in WTO Disputes." *International Organization* 68 (3): 663–99.

Johns, Leslie, and Krzysztof J. Pelc. 2016. "Fear of Crowds in WTO Disputes: Why Don't More Countries Participate?" *Journal of Politics* 78 (1): 88–104.

Johns, Leslie, and B. Peter Rosendorff. 2009. "Dispute Settlement, Compliance and Domestic Politics." In James C. Hartigan, ed., *Trade Disputes and the Dispute Settlement Understanding of the WTO: An Interdisciplinary Assessment.* Bingley: Emerald, 139–63.

Kelemen, R. Daniel. 2004. *The Rules of Federalism: Institutions and Regulatory Politics in the EU and Beyond.* Cambridge, MA: Harvard University Press.

Kelemen, R. Daniel. 2011. *Eurolegalism: The Transformation of Law and Regulation in the European Union.* Cambridge, MA: Harvard University Press.

Kim, Moonhawk. 2012. "Disguised Protectionism and Linkages to the GATT/WTO." *World Politics* 64 (3): 426–75.

Kono, Daniel Y. 2006. "Optimal Obfuscation: Democracy and Trade Policy Transparency." *American Political Science Review* 100 (3): 369–84.

Kucik, Jeffrey, and Krzysztof J. Pelc. 2013. "Measuring the Cost of Privacy: A Look at the Distributional Effects of Private Bargaining." *British Journal of Political Science* 46 (4): 861–89.

Kucik, Jeffrey, and Krzysztof J. Pelc. 2016. "Do International Rulings Have Spillover Effects? The View from Financial Markets." *World Politics* 68 (4): 713–51.

Locke, Richard M. 2002. "The Promise and Perils of Globalization: The Case of Nike." Working paper no. 02-008, MIT Industrial Performance Center, Cambridge, MA.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Pelc, Krzysztof J. 2009. "Seeking Escape: The Use of Escape Clauses in International Trade Agreements." *International Studies Quarterly* 53 (2): 349–68.

Pelc, Krzysztof J. 2014. "The Politics of Precedent in International Law: A Social Network Application." *American Political Science Review* 108 (3): 547–64.

Pelc, Krzysztof J. 2017. "What Explains the Low Success Rate of Investor-State Disputes?" *International Organization* 71 (3): 559–83.

Sattler, Thomas, and Thomas Bernauer. 2011. "Gravitation or Discrimination? Determinants of Litigation in the World Trade Organisation." *European Journal of Political Research* 50 (2): 143–67.

Shaffer, Gregory C. 2003a. *Defending Interests: Public-Private Partnerships in WTO Litigation*. Washington, DC: Brookings Institution.

Shaffer, Gregory C. 2003b. "How to Make the WTO Dispute Settlement System Work for Developing Countries: Some Proactive Developing Country Strategies." Working paper no. 5, International Centre for Trade and Sustainable Development, Geneva.

Steinberg, Richard H. 1999. "The Prospects for Partnership: Overcoming Obstacles to Transatlantic Trade Policy Cooperation in Asia." In Richard H. Steinberg and Bruce Stokes, eds., *Partners or Competitors? The Prospects for U.S.-European Cooperation on Asian Trade*. Lanham, MD: Rowman & Littlefield, 213–50.