Who's Listening? Audiences, Alarms, and International Cooperation

Stephen Chaudoin

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Politics
Adviser: Helen Milner and Christina Davis

September 2012

© Copyright by Stephen Chaudoin, 2012.

All rights reserved.

Abstract

A large body of literature with a lengthy history argues that international institutions facilitate cooperation by providing information. Cooperation among nations is difficult without credible punishment for defectors, and information is key to detecting the occurrence and severity of those defections. Domestic audiences are thought to be a key source of punishment. This dissertation explains how variation in the preferences and political strength of domestic audiences condition the informational role of institutions. I develop a theory that shows how audience preferences and strength affect how audiences react to information about defections, how their reaction, in turn, affects member states' strategic decision over whether to transmit information, and how policymakers choose whether to cooperate in the shadow of potential punishment. I demonstrate this theory with evidence at both the macro and micro levels, both observational and experimental. At the macro level, I show how audience preferences and political strength affect the timing of World Trade Organization disputes against the United States. At the micro level, I conduct an original survey experiment that shows how audience preferences moderate the degree to which audiences punish defections. Taken together, the theory and empirical analysis advance our understanding of the promise and limitations of international institutions and agreements as independent forces for cooperation.

Acknowledgements

I owe a large debt of gratitude to my advisers at Princeton University, Christina Davis, Robert Keohane, John Londregan and Helen Milner. I also owe appreciation to my advisers while I was an undergraduate student at Emory University: Clifford Carrubba and Eric Reinhardt. Numerous other people have made helpful suggestions on each of the three chapters: Michael Aklin, Phillip Arena, Marc Busch, Terrence Chapman, Rex Douglass, Songying Fang, Jennifer Ghandi, Alexander Hirsch, Sarah Hummel, Kosuke Imai, Matthew Incantalupo, Johannes Karreth, Jeff Kucik, Shawn Ramirez, Kristopher Ramsay, Emily Ritter, Tom Scherer, Dustin Tingley, and Johannes Urpelainen each gave valuable advice. For chapter 2, I benefitted from feedback given by audiences at the 2011 Annual Meeting of the International Political Economy Society and International Studies Association and from the 2012 Conference on the Political Economy of International Organizations. For chapter 3, I benefitted from feedback given by audiences at the 2010 annual meeting of the International Political Economy Society, American Political Science Association, and Midwest Political Science Association. For chapter 4, I benefitted greatly from research support from the Niehaus Center for Global Governance.

To my parents, Steve and Jenne.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

International relations scholars have long argued that international institutions and agreements facilitate cooperation by providing information about member state behavior (Keohane, 1984; Milgrom, North and Weingast, 1990). The logic is simple: for nations to cooperate, there must be a credible mechanism to punish defections. Information about the occurrence and severity of defections is key to any punishment mechanism. Without the ability to detect defections, punishment is ineffective.

A large and growing body of literature focuses on the *domestic* sources of punishment faced by policymakers who defect from cooperation. A policymaker could face electoral backlash from voters, economic punishment from market actors, or decreased support from powerful interest groups. In a world without international institutions and agreements, such defections may go undetected, and therefore unpunished, because domestic audiences may not be able to perfectly monitor their policymaker's decisions. But in a world with information-providing institutions and agreements, defections are detected and punished. The prospect of such punishment deters policymakers from defecting in the first place. By providing an alarm that sounds when policymakers defect, institutions and agreements facilitate cooperation. This logic forms the foundation of many well-known explanations for how institutions and agreement facilitate cooperation, such as those based on credible commitments (Simmons, 2000, 2009; Simmons and Danner, 2010;

Bthe and Milner, 2008), audience costs(Tomz, 2007, 2008; Levendusky and Horowitz, 2012; Fearon, 1994), and myriad others (Mansfield, Milner and Rosendorff, 2000, 2002; Rosendorff, 2005; Elkins, Guzman and Simmons, 2006; Dai, 2002, 2007; Fang, 2008).

The launching point for this dissertation is two observations about the world that diverge from these well-known stories. First, existing explanations describe institutions as "trip-wires" that are triggered whenever member states step out of line with policies that diverge from their international institutions. But few institutions act as trip-wires that sound immediately after any and all defections. The alarm sounds after some defections but not others. Even in the case when the alarm sounds, it often sounds only after a lengthy delay. This is because the sounding of the alarm is not automatic. It is a strategic decision made by members of the institution or parties to the agreement.

Second, existing explanations focus on pro-cooperation domestic audiences who have the political strength to influence policymaker decisions. Yet the political economy facing policymakers consists of multiple domestic audiences, who may vary in their support for for cooperation and in their political strength. Some audiences are pro-cooperation while others are pro-defection. The balance of political power can favor one or the other, and can also vary over time.

The foundation of this dissertation is an understanding of the following components: audience reactions to institutional alarms, the strategic decision to sound the alarm, and policymaker decisions over cooperation. Variation in preferences and political strength affects how audiences react to hearing the alarm. Will they punish policymakers for defections from cooperation or reward them? Is this punishment a strong or weak inducement for the policymaker to cooperate or defect? This variation in turn affects the strategic decision over whether to sound the alarm after defections. Will sounding the alarm, and potentially triggering domestic punishment, cause the offending government to change their policies? Are these changes worth the costs of sounding the alarm? And finally, how does the possibility of the alarm being sounded affect a policymaker's initial decision to cooperate or defect? Under what conditions does the alarm succeed in deterring defections?

Each of the three chapters that follow are presented as a stand-alone article, but they are very tightly tied by the theme of understanding how audience preferences and political importance strengthen or moderate the ability of institutions and agreement to facilitate cooperation via the provision of information.

The first chapter uses an original game theoretic model that links audience reactions, the decision to sound the alarm, and policymaker decisions. The theory is generalizable to many institutional settings and many issue areas of international cooperation. The theory shows the limits and potential of existing alarm-based explanations, where domestic audiences punish policymakers when the alarm sounds. Specifically, I show the conditions under which this dynamic can arise endogenously. On the one hand, these conditions are a cause for optimism. One condition requires that international institutions perform a very simply role: they must provide a costly mechanism by which one member state to accuse another member of defecting. On the other hand, the effectiveness of this mechanism is constrained by the preferences and political strength of domestic audiences. When domestic audiences support cooperation, an information-providing institution can be a powerful force for cooperation. But when domestic audiences oppose cooperation, information-provision can be ineffective or even counterproductive for cooperation. Similarly, when pro-cooperation audiences are politically strong, information provision can be effective, but when policymakers do not care about these audiences, information does not necessarily facilitate cooperation.

The model developed in the theoretical chapter also generates several empirically testable predictions. The theory shows how the preferences and political strength of audiences strength affect whether and when one member state will sound the alarm against another member state. Since sounding the alarm is costly, member states are most likely to choose this option when domestic audiences in the targeted state support cooperation and policymakers care about these audiences' preferences. Strong pro-cooperation audiences also cause their policymakers to choose more cooperative policies, *ex ante*.

The second chapter tests these predictions at the “macro” level in the context of World Trade Organization (WTO) disputes against the United States. U.S. legislation creates processes under which U.S. firms can petition government bureaucracies for tariffs on imports of competing goods. Many of the disputes brought against the United States by its trading partners under the WTO’s Dispute Settlement Understanding have sounded the alarm against these alleged violations of WTO rules. However, there is significant variation in whether and when trading partners choose to sue the United States over these barriers. Often, trade partners wait months or years before sounding the alarm.

I test the predictions generated by my theory by modeling the timing of WTO disputes against this set of U.S. trade barriers. Using a Bayesian multinomial probit model of competing risks, I find results consistent with the theory. Trade partners are more likely to sue the United States when both (a) broader audiences are less protectionist and (b) as elections approach. U.S. trade barriers are morel likely to be targeted with WTO disputes during election years, when policymakers are more attuned to the preferences of broader constituencies, and during times of lower unemployment, which are associated with less support for protectionism. During election years with high unemployment, the United States’ trading partners are more likely to delay disputes until after the election.

The results also show how current debates between Realist and Institutionalist explanations for member state behavior are too stark. I show how Institutionalist explanations- like those pertaining to the informational role of institutions- and Realist explanations- like those pertaining to power and retaliation- *both* explain important patterns in U.S. trade policy. The results regarding the timing of WTO disputes against U.S. trade barriers are consistent with institutionalist or information-based explanations. But the United States is also more likely to unilaterally remove trade barriers against partners that have higher trade leverage or power over the United States, and vice versa. While Realist verses Institutionalist debates are often cast as either-or propositions, this need not be the case empirically, where mechanisms associated with both theories are supported.

The third chapter tests the theory at the “micro” level using an original survey experiment using online recruits from Amazon’s Mechanical Turk experiment. In this experiment, I described a scenario to the respondents in which a U.S. firm was facing increased competition from foreign imports. The U.S. president had to decide whether to impose import restrictions on foreign-produced goods. Respondents were then told that the president decided in favor of the restrictions, and respondents then answered whether they approved or disapproved of the president’s actions. The experiment consisted of randomly treating respondents with various arguments for and against import restrictions.

The goal of the experiment is to decompose the respondent’s approval of the presidents actions into two parts: a consistency effect and a policy effect. Consistency effects describe how a respondent’s approval of the president is lower when the president’s actions are inconsistent with previous commitments, as hypothesized by audience costs theories. Policy effects describe how a respondent’s approval of the president is based on whether the president’s action matches the respondent’s policy preferences. To tap into consistency effects, some respondents received an international law treatment, which told them that import tariffs violated a U.S. free trade agreement and would likely result in a WTO dispute against the U.S.

The results show that consistency effects as hypothesized by audience costs theories are indeed present. When told that the president’s policies are inconsistent with prior commitments, respondents are significantly less likely to approve of the president’s actions. However, these effects are significantly moderated by policy preferences. Respondents who support or oppose free trade, i.e. respondents with a preference over trade policy, consistency effects were minimal. Consistency effects only mattered for respondents with no opinion on trade policy. And even for these respondents, the effects of placebo treatments were comparable to the effects of consistency treatments. Consistent with the predictions of the theoretical model developed in the first chapter, citizens’ reactions to hearing an alarm indicating that policymakers have broken their international promises are tempered strongly by the citizen’s preferences over the policy itself.

Chapter 2

Theory: What Makes Commitments Credible?

Most international institutions lack independent enforcement capabilities. As a result, a large and growing body of literature argues that domestic audiences play a crucial role in imposing costs on governments who defect from their international agreements. International institutions, and legalized dispute settlement mechanisms in particular, help facilitate international cooperation, because these bodies transmit information about member state behavior to those domestic audiences. When a member state violates the agreement, the institution acts as a fire alarm that alerts domestic audiences of their government's misbehavior. Hearing this alarm, the audience punishes the offending government, imposing a cost on governments who do not comply with their international obligations. This threat of *ex post* punishment helps facilitate cooperation, *ex ante*. This dynamic is at the core of many broader theories of the effects of international institutions pertaining to a wide range of issue areas, such as those based on audience costs (Tomz, 2007). or credible commitments (Simmons, 2000; Simmons and Danner, 2010) and is particularly emphasized in theories of dispute settlement (Mansfield, Milner and Rosendorff, 2000, 2002; Rosendorff, 2005).

If international institutions play an important fire alarm role, then two puzzles arise. First, why is there significant variation in whether or not the alarm sounds after violations? Consider the context of tariff barriers and the World Trade Organization's (WTO) Dispute Settlement Understanding (DSU). The vast majority of WTO-illegal trade barriers erected by WTO members do not result in any alarm-sounding DSU litigation. The DSU is among the most vibrant and active international courts in existence, having heard 427 cases as of January 2, 2011. Yet, few would doubt that hundreds, if not thousands, of explicit tariff barriers and hidden non-tariff barriers have escaped DSU scrutiny.

Second, why is there significant variation in the timing of the alarm? Returning again to international trade, members often allow illegal and harmful trade barriers to remain in place for months, or even years, before they sound the alarm with DSU litigation. If the victim need only sound the alarm in order to mobilize domestic audiences against their government's policies, then why wait? Few, if any, international institutions act as fire alarms that alert domestic audiences to government misbehavior *immediately* after *all* violations of an agreement. If the information-providing role of institutions and the resulting noncompliance costs are key explanations for international cooperation, then understanding whether and when the alarm will sound is of first order importance.

A key part of the answer comes from asking: who's listening to the alarm? The preferences and political strength of the groups hearing the institutional alarm are critical features of the fire-alarm dynamic that are frequently omitted from existing explanations. Often, the domestic audiences hearing the alarm are assumed to be monolithic and static. The audience is often assumed to be in favor of punishing their government for violations and this punishment is assumed to be of consequence to the government. However, audiences vary along both dimensions. Audiences can vary in their preferences. Domestic audiences often actively support non-compliant government policies and they can vary in the intensity of their dislike of defections. Audiences also vary in their political strength and their ability to influence government policymaking. Cross-nationally, not all governments care equally about potential audience punishment, which has driven much of the research on regime type and audiences costs. But even within a particular regime, government sensitivity to audience punishment can vary over time, e.g. with the electoral cycle.

I develop a theory showing how audience features affect the ability of international institutions to generate noncompliance costs. These noncompliance costs are, in essence, what makes commitments credible and what makes audience-imposed punishment costly. The preferences and strength of uninformed audiences affect the magnitude of noncompliance costs, which affects governments' compliance and dispute decisions. The theory draws from existing work on domestic constitutional courts which argues that the anticipated reaction of domestic audiences affects the relationship between the court and other lawmaking branches of government (Vanberg, 2001, 2005; Carrubba, 2005, 2009; Staton, 2006). I also draw from existing work arguing that international courts and information provision facilitate international cooperation (Mansfield, Milner and Rosendorff, 2000; Carrubba, 2005; Carrubba, Gabel and Hankla, 2008; Rosendorff, 2005).

The key insight is that there is good news and bad news for existing theories of noncompliance costs. The good news is that there are very minimal requirements for a dynamic to arise in which institutions can generate noncompliance costs, i.e. for commitments to be credible and audience punishment to be costly. The institution need only provide a public and costly mechanism for governments to use as a signal to domestic audiences, and the preferences of the government sending the signal need only be partially aligned with those of the intended audience. The bad news is that the magnitude of these noncompliance costs, and therefore their ability to influence government behavior as argued by audience costs and credible commitment theories, is constrained by the preferences and strength of those audiences. The institution cannot facilitate cooperation beyond the level desired by the audience.

The theory also generates predictions for how audience features affect the behavior of governments pre- and post-alarm, as well as the probability that one government will choose to sound the alarm in the first place. Sounding the alarm is most valuable to the plaintiff country when domestic audiences in the defendant country are most "favorable," i.e. the audience prefers similar changes to the defendant government's policies as the plaintiff desires *and* when the defendant government cares about those audiences. Defendant governments engage in less severe violations of their agreements when they must make policy in the shadow of disputes that could potentially activate such audiences.

I demonstrate the theory empirically with evidence from the WTO's DSU. Audience features explain why some DSU disputes succeed in mobilizing audiences against protectionist barriers, and why other disputes backfire. The costliness of DSU disputes is also a key explanation for why DSU disputes play a greater role in providing information than do other forums. The theory also explains a troubling empirical puzzle: democracies are often thought to have lower tariff barriers (Milner and Kubota, 2005), yet they are far and away the most frequent DSU defendants. I argue that this is because even if democracies are "better behaved" with regards to their agreements, they are also the best targets for audience-mobilizing disputes, which is supported by cross-national analysis of dispute frequencies. Though I use examples from international trade, the argument is general- describing international institutions where governments choose compliance policy in the shadow of uninformed audiences who can potentially learn from a costly, institutional signalling mechanism.

The next section reviews the relevant theoretical literature on dispute settlement and information transmission. The third section describes the model, its results, and supporting empirical evidence. The fourth section concludes and also discusses the model in the context of a prominent normative debate over whether international institutions enhance or hinder democracy.

Cooperation, Courts and Information

International cooperation is often thought of as a prisoner's dilemma-style interaction among governments. Governments can potentially benefit from cooperating by making mutual policy adjustments (Keohane, 1984). But the costliness of these adjustments make defecting from cooperation tempting. International institutions are thought to facilitate cooperation by making it more costly for governments to defect, and these costs are often called noncompliance costs.

Since most institutions lack independent enforcement powers, many theories, such as those based on credible commitments or audience costs, examine domestic sources of noncompliance costs.[1] In audience costs explanations, the audience is often thought of as a set of voters who care about their government's policy choices. Noncompliance costs arise as the result of electoral punishment: voters punish their elected officials for breaking their agreements by not returning them to office. As Tomz (2007) argues, audience costs are "the surge in disapproval that would occur if a leader made commitments and did not follow through" (pg. 823). Credible commitments theories are founded on a similar dynamic. In her theory about the effects of IMF obligations, Simmons (2000) argues that a government's Article VIII obligations "mobilizes a new set of external actors (private economic, governmental, and legal) who may exert pressure to comply on a government that is considering or engaging in rule violation" (pg. 821). These types of arguments have been made in a variety of contexts, ranging from international trade agreements (Bthe and Milner, 2008; Mansfield, Milner and Rosendorff, 2000) to bilateral investment treaties (Elkins, Guzman and Simmons, 2006) to human rights (Simmons, 2009) to war crimes (Simmons and Danner, 2010).

International institutions are crucial to these theories because they provide information about government behavior to otherwise uninformed audiences (Keohane, 1984; Milgrom, North and Weingast, 1990). The audiences who potentially impose noncompliance costs often cannot perfectly monitor government behavior: a voter may not know whether their government has erected illegal trade barriers; a private investor may not be certain about whether a potential host government is likely to expropriate their investments. If they do not know whether or not a government has misbehaved, then audiences cannot use the threat of punishment to incentivize governments to

[1] A related explanation considers retaliation by other member states, often in a repeated-play setting. Here, I focus on domestic noncompliance costs as opposed to costs incurred when a defection triggers punishment from other member states.

comply. Institutions alleviate this problem by acting as alarms that alert uninformed audiences to government noncompliance (Mansfield, Milner and Rosendorff, 2000).[2]

Institutions often, and increasingly, include judicial dispute settlement mechanisms that have been prominently linked to information transmission in a number of theories (Mansfield, Milner and Rosendorff, 2000; Carrubba, 2005; Carrubba, Gabel and Hankla, 2008; Rosendorff, 2005). In the context of the World Trade Organization and its Dispute Settlement Understanding, B. Peter Rosendorff argues that

> [Dispute settlement] serves a crucial information-providing role. It establishes the facts, adjudicates on a violation, estimates the damages, and reports a successful completion of the process. It is this informational role of the [DSU] that determines its effectiveness in the world trading system. (2005, pg. 391)

Institutions, and dispute settlement mechanisms in particular, therefore help ameliorate monitoring problems by establishing an information-providing alarm that sounds whenever a government defects from its agreement. The institution facilitates international cooperation because noncompliance is costlier in a world with an alarm, where audiences learn about and punish defections, than in a world without an alarm, where governments are left to their own devices.

However, the audiences in these theories are often assumed to have two features: they support compliance by their government and they have the capacity to impose costs on their government when it defects. In reality, audiences vary significantly along both dimensions.[3] With regards to audience preferences, audiences do not always support policies that are consistent with their government's international obligations, and often support defections from agreements. In the case

[2] Note that the assumption made here, and in the theory that I develop, is very minimal and general: I do not assume that institutions have more/private information over government behavior or that they are highly legalized or sophisticated. I only assume that institutions can provide a simple, binary piece of information, namely, they are forums in which an alarm can sound or remain silent.

[3] For two notable exceptions, see Rickard (2010) and Tomz and Van Houweling (2012). Rickard analyzes how different electoral systems amongst democracies and the preferences of their constituents affect compliance behavior. Tomz and Van Houweling analyze survey responses to scenarios in which candidates switch positions, accounting for the respondent's policy preferences.

of trade and the WTO, domestic political audiences often support protectionist measures and oppose compliance with adverse WTO rulings. Support for free trade can vary across individuals (Mansfield and Mutz, 2009; Hays, Ehrlich and Peinhardt, 2005). Support for free trade can also vary across time, waxing or waning depending on macroeconomic conditions (Bergsten and Cline, 1983). Similar variation is likely in every other context in which audience costs/credible commitments arguments are made. In the context of investment, domestic constituents may vary in their support of a government that expropriates foreign assets, and foreign investors may vary in the degree to which they fear expropriation. In the context of human rights, constituents in one country may vary in the degree to which they demand that their government address human rights violations in other countries.[4]

Audiences also vary in their ability to inflict punishment on their government. Governments vary in the degree to which they care about the preferences of broad audiences relative to specialized interest groups (Gawande, Krishna and Olarreaga, 2009). Cross-national variation in the degree to which governments care about audience preferences has often been linked to regime type, with democracies thought to care more about audiences than non-democracies.[5] Government sensitivity to audience preferences also varies temporally. In the run-up to democratic elections, politicians are particularly attuned to the preferences of their constituents. Canes-Wrone and Shotts (2004) argue that variation in presidential approval ratings can affect the responsiveness of executives to public preferences. Dai (2007) considers how interests groups with different, exogenously generated monitoring abilities can influence the behavior of politicians. Interest groups with the greater monitoring capacity, i.e. groups who can better discern the government's policies from "noise," have greater influence on government policy.

[4] Snyder and Borghard (2011)'s recent criticism of audience costs arguments in the context of crisis-bargaining questions the assumption that audiences care about policy consistency apart from their preferences over policy substance.

[5] This is the main focus of audience costs arguments in the context of security and crisis bargaining (Fearon, 1994). For important exceptions, see Slantchev (2006) and Weeks (2008).

I argue that variation in the preferences and strength of audiences should have a significant effect on virtually every aspect of audience costs and credible commitments theories. To generate intuition on why this might be the case, consider a related literature on domestic constitutional courts, which is keenly focused on audience features and how they affect different aspects of judicial behavior. Like most international institutions, domestic courts lack often lack independent enforcement power. How then, can domestic courts constrain policymakers who might otherwise be free to ignore their rulings? The answer for many domestic courts scholars is based on the audiences who observe those rulings (Vanberg, 2001, 2005; Carrubba, 2005, 2009; Staton, 2006). As Georg Vanberg (2005) writes:

> ... the interactions between courts and other policymakers do not occur in a vacuum... If citizens value judicial independence and regard respect for judicial rulings as important, a decision by a elected official to resist a judicial ruling may result in a loss of public support... The fear of such a backlash can be a forceful inducement to implement judicial decisions faithfully (20).

A key insight of the domestic courts literature is that audience features affect judicial behavior. If the audience does not support adherence to a particular judicial ruling or if the informational setting is such that audiences are unlikely to learn about policymaker disobedience even when it does result in judicial scrutiny, then policymakers are more free to choose policies to their liking and courts are less likely to rule against those policies (Vanberg, 2001, 2005). Domestic courts strategically publicize important rulings, based on the anticipated reaction of public audiences (Staton, 2006). Carrubba (2005) applies a similar model to an international cooperation setting, showing how an institutional mechanism that reveals the costs to a member state of noncompliance can help governments coordinate their punishment strategies so as to punish governments for low-cost defections from and agreement and forgive governments for high-cost defections.

Audience features should similarly affect government behavior in the international cooperation settings considered by credible commitments and audience costs theories. Audience features affect how the audience reacts to learning about its government policies, once the institutional alarm is sounded. A compliance-supporting audience might react negatively to learning that its government has broken its international obligations, while a noncompliance-supporting audience might react with ambivalence or even support for further noncompliance. A government facing backlash from politically strong pro-compliance group, might be less inclined to defect in the first place, while a government facing a weak backlash might be less fearful of the repercussions from defections.

Audience features also affect the decision over whether or not to use an international institution to transmit information in the first place. In the case of dispute settlement mechanisms, information about noncompliance is only transmitted when one government makes the strategic decision to bring litigation before the institution's judicial body. The sounding of the alarm is rarely automatic. Recent research has begun to consider how noncompliance costs affect the litigation decisions of governments in international cooperation settings. For example, Songying Fang (2010) and Michael Gilligan, Leslie Johns, and Peter Rosendorff (2010) develop models that focus on the effects of institutional "strength" on the occurrence of disputes. Two countries bilaterally negotiate over an issue and have the option of appealing to an international dispute settlement body for a ruling over that particular issue. Gilligan, Johns and Rosendorff (2010) emphasize how variation in the noncompliance costs imposed on a member state who disobeys the institution's ruling varies across-institutions and how this affects disputes. Fang (2010) emphasizes how variation in these costs across countries affects disputes.[6] Building on this literature with *exogenous* noncompliance costs, I consider how noncompliance costs might arise *endogenously* and how features of the audience imposing those noncompliance costs affect the decision over whether to sound the alarm in the first place.

[6] Several related theories consider the coordinating role of international judicial bodies. Carrubba (2005) argues that courts can help facilitate cooperation by revealing the costs of compliance. Johns (Forthcoming) describes how disputes can transmit information and trigger punishment by third parties, such as domestic political actors. The costliness of initiating a dispute facilitates a screening mechanism whereby member states can use dispute settlement to coordinate enforcement of the institution's judicial decisions by "disinterested" third parties.

This theory is also related to arguments describing the signalling role of appeals to international institutions (Chapman, 2007; Fang, 2008; Chapman, 2009; Chapman and Wolford, 2010). In these arguments, a politician proposes a policy and then chooses whether or not to appeal to an international institution. When an appeal is made, another actor (such as a foreign government or the pivotal voting member of that institution) gets to send a signal to the politician's domestic audience indicating approval or disapproval of the politician's policy. Upon observing that signal (if an appeal is made), the audience can update their beliefs about the quality of the policy being proposed and can punish or reward their politician accordingly. A common result across the models is that the relationship between the audience's preferences and the signal sender's preferences affect both the signal sent and likelihood that a politician will choose to appeal to the institution in the first place. In these models, the politician or government who chooses a policy is the ultimate gatekeeper over whether or not a signal pertaining to the effects of that policy is sent to a domestic audience. In order to better match institutional settings with dispute settlement mechanisms, the model described here differs in one key way: another government chooses whether or not to initiate a dispute, and thus send a signal of noncompliance. In a judicial setting, the government being accused of noncompliance cannot "decline" the charges levied by the accuser. The government choosing its compliance policy does not have perfect control over whether or not any dispute signal will be sent- that decision is ultimately made by other member states who are potentially harmed by noncompliance.

On a final note, to be sure, information transmission is not the only role of international institutions and their legalized dispute settlement mechanisms. Christina Davis (2011) argues that governments can use disputes to reassure domestic groups that the government is committed to defending their interests. Chad Bown (2005*b*) finds that trade barriers are more likely to be challenged at the WTO when the stakes of the case are higher, and when the defendant country does not have a similar retaliatory mechanism. Allee and Huth (2006) analyze when countries choose legalized dispute, rather than bilateral negotiation, as a way to settle territorial disputes. They argue that legalized dispute settlement provides political cover for policymakers, and empirically, countries with stronger domestic opposition groups and more democratic dyads are more likely to pursue legalized dispute settlement. It is worth emphasizing that these arguments for why countries choose legalized dispute settlement are not mutually exclusive with the information transmission mechanism described here. No one theory completely explains all of the incentives and constraints facing governments contemplating legalized dispute settlement.

The Model

Two countries are trading partners and are members of an agreement that allows them to initiate costly disputes over tariff policies. There are the three players in the model: the government of the "Home" country, $Home$, the "Foreign" government, $Foreign$, and an $Audience$ within the home country. Each player cares about the tariffs, $t \in \Re$, that the home government levies against imports from the foreign country. The audience can be thought of as any group that lacks perfect information about the home government's tariff policies. For instance, "downstream" firms paying inflated prices for intermediate production materials may lack perfect information about the tariff policies responsible for those higher prices. Consumers are even more uninformed about these policies. These audiences can potentially engage in some costly action to try and influence the home government's policies. For instance, firms could pay the costs associated with mobilizing into an organized interest group, or constituents can mobilize to punish elected officials, as in the familiar audience costs argument.

Each of the three players has preferences over the tariff set by the home government.[7] The foreign government prefers lower tariffs, and its preferences over tariffs are represented by the utility function: $u_F(t) = -t$. The audience has a most preferred tariff level, $t = A$, and its

[7] In some models, like that of Mansfield, Milner and Rosendorff (2000), preferences over tariff levels are generated by an underlying economic model. Groups with different factor endowments or technologies have different preferences over tariffs as a result of the economy or market in which they will operate. For simplicity, I leave the microfoundations of these preferences unspecified, but their existence and the potential for preferences to diverge across groups is well established elsewhere.

preferences over tariff policy are represented by the function: $u_A(t)$, which is maximized at $t = A$, concave, decreasing in t when $t > A$, and increasing in t when $t < A$.[8]

The home government's most preferred tariff policy, H, depends on its type. The home government can be a "good" government from the perspective of the audience, and have preferences identical to those of the audience, where $H = A$. Alternatively, the home government can be a "bad" type whose most preferred policy is $t = B > A$.[9] The preferences of the home government are represented by $u_H(t)$ and have the same properties as the audience's utility function, apart from the point at which the function is maximized. The probability of a bad home government, $Pr(H = B)$, is $\lambda \in (0, 1)$ and is commonly known. The audience does not observe their government's type.

The sequence of the game is as follows. First, Nature selects the home government's type. Next, the home government chooses their initial tariff level, t_1. The foreign government observes the home government's type, their initial policy, and draws the costs to initiating a dispute, k, from a commonly known distribution, $F(k)$, which is uniform on the interval $[\underline{k}, \overline{k}]$, with $\underline{k} < 0 < \overline{k}$.[10] The foreign government then chooses whether or not to initiate a dispute, D or $\sim D$.

The audience observes the foreign government's decision over whether to initiate a dispute and then decides whether to pay costs, $m > 0$, and mobilize to influence the policy chosen by the home government. If the audience chooses not to mobilize, $\sim M$, then the initial policy chosen by the home government, t_1, is the final policy. If the audience chooses to mobilize, M, then the home government chooses a new policy, t_2, and must partially internalize the preferences of their audience. Specifically, the home government must choose their post-mobilization final policy

[8] I describe a single audience as opposed to a collection of audiences for simplicity. The preferences of the audience could also be thought of as an aggregation of the preferences that arises in a common agency setting, like that of Bernheim and Whinston (1986) or Grossman and Helpman (1994).

[9] There are many ways that politics can drive a wedge between the preferences of the government and the preferences of a particular audience. Mansfield, Milner and Rosendorff (2002) use a fully specified economy to generate preferences over tariff policy. Grossman and Helpman (1994) model government preferences as an aggregation of concern for social welfare and special interest group contributions.

[10] Whether or not the foreign government observes the home government's type does not affect results. The foreign government only cares about the home government's type insofar as it affects the home government's policies. To condense notation, I will refer to $F(k)$ and $f(k)$ as the distribution and accompanying density function.

by maximizing an α-weighted combination of their own preferences and those of the audience: $U_H(t_2) = \alpha u_A(t_2) + (1 - \alpha)u_H(t_2)$.[11]

The decision to mobilize can be thought of as a decision to gather precise information about the home government's policy, mobilize politically to lobby the government, or make political contributions that are conditioned on changes to policy. All of these are costly actions that can make the home government pay more attention to the preferences of that audience. $\alpha \in [0, 1]$ represents how much the home government cares about the audience, should the audience mobilize. For example, if $\alpha = 1$, mobilization causes the home government to act as though it were a member of that group. If $\alpha = 0$, mobilization has no effect. The audience does not observe the initial policies chosen by the home government or the home government's type, but can potentially condition their mobilization decision on whether or not the foreign government initiates litigation.

For concreteness, I describe the model in terms of tariffs and international trade, but the model is much more general. t could be thought of as any policy covered by an international agreement, where governments can choose policies that are more or less in compliance with their obligations. In pollution control agreements, governments can comply by meeting their abatement targets, or retain higher levels of pollution than allowed. In investment agreements, governments can choose expropriation policies, like tax breaks for domestic firms, that are more or less harmful to foreign firms. Governments can choose to respect human rights, or they can engage in human rights violations of varying degree and severity. In many contexts, relevant audiences lack information about these policies, and other governments have recourse to costly dispute settlement mechanisms.

Information Transmission Equilibrium

The tension that arises in the model is similar to the concept of agency slack. The audience is akin to a principal, who would like their agent, the home government, to choose policies in line with

[11] This assumption is a reduced form of an electoral or political constraint. In the common agency settings mentioned above, the equilibrium policy chosen more heavily "weights" the interests of mobilized groups. The assumption made here simply says that after mobilization, the government must assign more weight to that group's preferences.

the principal's preferences. But the potential divergence in preferences between the principal and agent, combined with the principal's inability to observe the agent's actions, allows the agent to choose policies that stray from the desires of the principal. This model examines the conditions under which a third party, in this case- the foreign government- who has preferences that are partially aligned with those of the principal, can strategically use costly disputes as signalling mechanism that enhances the principal's control over their agent.

I first establish the conditions under which an "information transmission equilibrium" (ITE) exists. An ITE has the features that are associated with information transmission or audience costs or credible commitments in the literature. A government signs an agreement, and if they violate the agreement and an institutional alarm sounds or a dispute occurs, then that government suffers some additional noncompliance costs or punishment. In this model, an ITE is one in which the foreign government's decision to initiate a dispute causes the home audience to mobilize with the goal of changing policy. Without the dispute, the audience does not mobilize. In other words, audiences condition their behavior on the signal sent by an institution or dispute.

Proposition 1. *There exists an information transmission equilibrium, such that,*

- *The audience chooses* $M|D$ *and* $\sim M| \sim D$
- *The foreign government chooses* L *if* $t_1 - t_2^* \leq k$
- *Good home governments choose* $t_1^* = A$ *and* $t_2^* = A$
- *Bad home governments choose* $t_1^* \in (A, B)$ *and* $t_2^* \in (A, t_1^*)$
  - *The probability of* D *for a good government is* $F(0)$
  - *The probability of* D *for a bad government is* $F(t_1 - t_2^*)$

if and only if:

(i) $Pr(H = B| \sim D)[u_A(t_{2b}^*) - u_A(t_{1b}^*)] \leq m \leq Pr(H = B|D)[u_A(t_{2b}^*) - u_A(t_{1b}^*)]$

(ii) $Pr(H = B|D) > Pr(H = B| \sim D) > 0$.

Proof. For the audience to choose $M|D$, it must be the case that $EU_A(M)|D \geq EU_A(\sim M)|D$. I call the optimal initial policies chosen by bad governments t^*_{1b} and t^*_{2b}. Rewriting the expected utilities:

$$Pr(H = A|D)u_A(A) + Pr(H = B|D)u_A(t^*_{2b}) - m \geq Pr(H = A|D)u_A(A) + Pr(H = B|D)u_A(t^*_{1b})$$

$$m \leq Pr(H = B|D)[u_A(t^*_{2b}) - u_A(t^*_{1b})]$$

where $Pr(H = B|D) = \frac{\lambda F(t^*_{1b} - t^*_{2b})}{\lambda F(t^*_{1b} - t^*_{2b}) + (1-\lambda)F(0)}$.

For the audience to choose $\sim M| \sim D$, $U_A(\sim M)| \sim D \geq U_A(M)| \sim D$..

$Pr(H = A| \sim D)u_A(A) + Pr(H = B| \sim D)u_A(t^*_{1b}) \geq Pr(H = A| \sim D)u_A(A) + Pr(H = B| \sim D)u_A(t^*_{2b}) - m$

$m \geq Pr(H = B| \sim D)[u_A(t^*_{2b}) - u_A(t^*_{1b})]$

where $Pr(H = B| \sim D) = \frac{\lambda[1 - F(t^*_{1b} - t^*_{2b})]}{\lambda[1 - F(t^*_{1b} - t^*_{2b})] + (1-\lambda)[1 - F(0)]}$.

The remaining parts of the proof are developed in subsequent propositions. □

Condition (i) of Proposition 1 says that mobilization costs must be high enough to keep the audience from always mobilizing and low enough to allow them to mobilize when they observe a dispute. If mobilization costs were very low, then the audience would want to mobilize even in the absence of a dispute, causing the foreign government to always eschew disputes, since they don't gain any additional benefits from a dispute. If mobilization costs were very high, the audience would not want to mobilize, even after observing a dispute, again causing the foreign government to avoid disputes.

Condition (ii) is straightforward in terms of the intuition of signalling models, but counterintuitive in its implications for the role of litigation costs in international dispute settlement. Condition

(ii) says that the audience's posterior belief about probability that their government is bad has to be higher after observing a dispute than in the absence of a dispute. The signal, i.e. the dispute, that the audience receives has this effect because litigation is costly, and therefore informative, to the audience. If litigation costs were too low, then the audience would not gain enough information from the signal to justify spending mobilization costs. The optimal level of litigation costs, from the audience's perspective, is not zero. If the audience could pick the distribution of litigation costs, they would balance two concerns: on the one hand, they want the signal to be sent often, but on the other hand, they want the signal to be withheld frequently enough so that it retains its informative value.

The costliness of different dispute settlement institutions affects the degree of scrutiny that government policies received from disputes, and why some dispute settlement bodies have much higher profiles than others. In 1999, Chile increased tariffs on vegetable oils from Argentina which had a significant effect on Argentine vegetable oil exports to Chile. Argentina first tried to address the tariffs bilaterally, and then through MERCUSOR's dispute settlement system. Chile refused to adjust the tariffs, and even strengthened them. Argentina then took Chile to the WTO's dispute settlement mechanism in 2000. Describing Argentina's experience with regional dispute settlement, Tussie and Delich (2005) observe that "The [MERCUSOR] dispute system was out of the public eye and at the same time it was both fast and low-cost. Chile did not, meanwhile, modify its reclassification." In contrast, their description of Argentina's experience with the WTO's dispute settlement mechanism notes both the costliness and additional exposure gained from the WTO's mechanism relative to MERCUSOR's:

> Although accessible only to highly profitable sectors because participation is too costly and time consuming, the WTO provides the intangible benefit of exposure. Pressure through exposure can help countries unable or unwilling to retaliate to obtain more favourable results than in bilateral or regional instances.

The existence of an information transmission equilibrium also requires the partial alignment of preferences between the foreign government and the audience. For tariff policies that are greater than the audience's ideal point, the foreign government and the audience both prefer lower tariffs than the home government. But if the audience preferred higher tariffs than the government, then the information transmission dynamic breaks down. If the audience preferred higher tariffs than the government, and disputes caused those audiences to mobilize, then the foreign government would not want to ever initiate disputes for fear of activating a protectionist audience. In such a case, the foreign government would only file disputes when they drew sufficiently negative litigation costs to offset the worsening of policy that resulted from the dispute. Snyder and Borghard (2011)'s recent critique of the theory of audience costs in the context of crisis bargaining notes how the omission of audience preferences in most theories of audience costs is important, because of the possibility that the public has *more* hawkish or dovish preferences than their political leaders, and that this divergence implies that audience costs need not always be present.

An example of dispute settlement inadvertently activating an extreme audience arose in a WTO dispute between Japan and the European Communities as complainants and Canada as the respondent. In 1965, Canada and the United States signed a bilateral agreement that lowered tariffs on trade in the auto industry between the United States and Canada. Approximately four years after the entry into force of the new WTO regime, in 1994, Japan and the European Communities challenged U.S. Canada auto agreement at the WTO's new dispute settlement body on the grounds that the pact violated the WTO's Most Favored Nation (MFN) rules against providing special treatment to only select trading partners. The auto pact with the United States was very popular in Canada and credited with generating significant economic growth, and was supported strongly by interest groups representing the auto sector. As a result, the audiences activated by the WTO dispute proved extremely hostile to changing this policy in the way desired by the complainants. According to one observer:

> Despite facing almost certain defeat, Canada vigorously defended and then appealed on the matter at the WTO. ... there was considerable public pressure on federal officials to take a strong stand not only in favour of the cherished Auto Pact but also against ’interference’ by an international body on a matter of domestic public policy. Once the WTO claim was made public, the significant media attention and the corresponding ’court of public opinion’ limited the government’s ability to enter into a negotiated settlement. At that point, the government had virtually no choice but to defend the Auto Pact vigorously even in the face of certain defeat (Krikorian (2005)).

Ironically, the end result of the WTO dispute was for Canada to *raise* its tariffs, applying them to more countries, in order to comply with MFN rules. The ability of dispute settlement to activate domestic audiences is not always a force for increasing the amount of international cooperation associated with an international institution.[12]

Effects of Audience Features on Equilibrium Behavior

The second set of questions motivating the model concerns how audience features affect government behavior. First, consider the effects of audience features on post-dispute policy. If disputes can trigger audience mobilization, then how would mobilization affect the home government’s updated policy, t_2^*? After mobilization, the home government faces the following optimization problem:

$$max_{t_2} \;\; \alpha u_A(t_2) + (1-\alpha) u_H(t_2)$$

Proposition 2. *The optimal post-mobilization policy, t_2^* satisfies:* $\frac{\alpha}{1-\alpha} = \frac{u'_H(t_2^*)}{-u'_A(t_2^*)}$.

[12] In a separate context, a similar dynamic arose during negotiations over the transfer of the Panama Canal in the 1970’s. In her study of the effects of secrecy on interest group activities, Barbara Koremenos (2012) describes how increased information about negotiations between the United States and Panama (via Panamanian government leaks) had the effect of activating previously laten interest groups who opposed the treaty. This ultimately made ratification more difficult.

Proof. The proof follows from the first order conditions of the post-mobilization maximization problem, $\alpha u'_A(t_2^*) + (1-\alpha)u'_H(t_2^*) = 0$. □

Proposition 2 says that, the ratio of the audience and home government's marginal utilities matches the (inverse) ratio of their strength after mobilization. If the home government and audience's utility functions, u_H and u_A, were identical apart from their maximization points and were symmetrical, then the optimal policy would be an α-weighted combination of the two ideal points, $t_2^* = \alpha A + (1-\alpha)H$.[13] If the audience and the home government share the same ideal point, $A = H$, as in the case of a "good" government, then $t_2^* = A$.

Corrollary 1. *In equilibrium:*
(i) $\frac{\partial t_2^*}{\partial A} > 0$, *(ii)* $\frac{\partial t_2^*}{\partial \alpha} < 0$, *and (iii)* $\frac{\partial t_2^*}{\partial B} > 0$, *for bad home governments.*

Corollary 1 and Figure 2.1 show how audience features affect the optimal post-mobilization policy. As the audience and the home government prefer higher tariffs, the home government will choose higher tariffs after mobilization.[14] As the audience's strength increases, the optimal policy decreases. Stronger audiences "pull" the optimal policy downward, with greater weight, towards the ideal policy of the audience.[15]

Figure 2.2 shows how the effects of audience preferences on policy are conditioned by the audience's strength. For example, the effects of a change in audience preferences can be magnified by the audience's strength, when the audience is stronger. A marginal increase in the audience's ideal tariff will have a larger effect on the final policy when α is higher than when α is lower. On the other hand, when α is low, or zero, changes in audience preferences have dampened effects on the final policy, or no effect at all. From the above example of symmetric utility functions, where $t_2^* = \alpha A + (1-\alpha)H$, the derivative of t_2^* with respect to A is simply α.

[13] For instance, this would be the case if both the home government and audience held preferences represented by the familiar quadratic loss function.

[14] From Proposition 2, for a fixed α, increasing A means that u'_A increases by the concavity of u_A, so u'_H must increase, which means a higher t_2^* by the concavity of u_H. The same argument applies for increases in H.

[15] Increasing α means $u'_H(t_2^*)$ must increase and $u'_A(t_2^*)$ must decrease, implying that t_2^* must increase.

These empirical findings of Dai (2007) are consistent with this conditional effect of audience preferences and strength. When pro-compliance interests groups compete with anti-compliance interest groups, the policy chosen by the government is more compliant as the electoral leverage and monitoring ability of the pro-compliance interest group increases. Analyzing the 1985 Sulfur Protocol of the LRTAP convention, she finds that countries with pro-compliance (pro sulfur-reduction) interest groups that were politically stronger and better able to monitor their governments enacted policies that resulted in greater reductions in sulfur emissions.

The foreign government chooses to initiate a dispute when the benefits outweigh the costs. The foreign government potentially benefits from a dispute if a dispute causes the audience to mobilize, and thus change the home government's policy from its initial tariff, t_1, to a new policy, t_2. In an information-transmission equilibrium, audiences mobilize only after disputes. The utility to the foreign government of initiating a dispute is $-t_2^* - k$, and the utility of not doing so is $-t_1$. Recall, for a good home government, $t_2^* = A$, and for a bad home government, $t_2^* > A$. In an information-transmission equilibrium, the foreign government initiates a dispute if and only if their costs are lower than their expected gains:

$$k \leq t_1 - t_2^*$$

For a good home government, therefore, the foreign government only initiates a dispute if it draws a negative litigation costs, i.e. it has some extraneous benefit to initiating a dispute, apart from the potential effects on home's policies.[16] Facing a bad home government, the benefit of a dispute comes from the effect that any subsequent audience mobilization will have on changing the initial tariff policy to a new, lower final policy. If the foreign government draws a litigation cost that is higher than the benefits from changing the home government's policy, then it will not initiate a dispute. The probability of a dispute for a particular initial policy, which I call $\Pi(t_1)$, is

[16] For instance, Davis (2011) argues that some countries initiate WTO disputes as a way to placate domestic industries.

the probability that the foreign government draws a low enough litigation cost that it will choose to initiate a dispute.

$$\Pi(t_1) = Pr(k \leq t_1 - t_2^*) = F(t_1 - t_2^*)$$

For a particular initial policy, features of the audience have a straightforward effect on the probability of a dispute. As the audience prefers lower tariffs, the expected gains from mobilizing that audience with a dispute increase, which increases the probability of a dispute by expanding the range of litigation costs over which the foreign government's gains outweigh their costs. As the audience grows stronger, the benefits from a dispute also increase, increasing the probability that the foreign government will draw litigation costs low enough to justify a dispute.

Proposition 3. *For a fixed initial tariff, t_1, and, when $H > A$, the probability of a dispute, $\Pi(t_1)$, is: (i) decreasing in A, (ii) increasing in α, and (iii) decreasing in H.*

Proposition 3 shows how features of the audience affect the foreign government's cost-benefit calculations for a dispute. The ideal audience for the foreign government to mobilize with a dispute is one that prefers lower tariffs and which has more sway over their government's policies. Audiences that prefer higher tariffs do not make attractive allies for the foreign government. Similarly, impotent audiences are not worth paying litigation costs to activate. As the home government prefers higher tariffs, it will be more recalcitrant in the face of a mobilized audience, which makes disputes less attractive.

These results explain a puzzling contradiction in the empirical evidence on tariffs and legalized WTO disputes. Milner and Kubota (2005) argue that, among less developed countries, democracies have lower tariffs, and Mansfield, Milner and Rosendorff (2002) argue that democracies are more likely to sign trade agreements with each other. Yet, despite their apparent penchant for lower tariffs, the most frequent respondents (defendants) in WTO disputes, by far, are established democracies: the United States and the European Union. Autocracies are very rarely defendants in WTO disputes. If democracies are so free-trade-loving and have lower tariffs, then why do they find themselves in court over illegal trade barriers so often?

The model's results argue that countries who are more likely to respond to disputes with lower tariffs make themselves more attractive targets for disputes. If democracies are more likely to have constituencies that prefer lower tariffs and are more sensitive to the preferences of these audiences in general, then it is not surprising that they are frequent defendants. Governments often erect tariff barriers in response to the protectionist pressures of concentrated groups, and to the detriment of the welfare of broader, more poorly informed constituencies. If the government cares less about potential backlash from these broader constituencies, then a dispute will not cause the government to change its tariff policy substantially. If the government cares more about this backlash, then they are more likely to change their policies after a dispute, which makes a dispute more attractive for the foreign government.

Empirically, countries that are more sensitive to the aggregate preferences of their constituents are targeted with more WTO disputes. Gawande, Krishna and Olarreaga (2009) use trade data to estimate a weighting parameter that measures the concern for aggregate welfare displayed by a country's leaders.[17] They obtain an estimate for this parameter (also called α) where higher values indicate that a government cares more about aggregate welfare in choosing its policies, as opposed to special interest groups. According to the model here, governments with higher sensitivity to audience preferences are more attractive litigation targets.

Figure 2.3 plots the number of WTO disputes against a country verses the Gawande, Krishna and Olarreaga (2009) measure of government sensitivity. The data support the model's prediction. Even excluding the United States and European Union, more sensitive governments are also more frequent WTO respondents.[18] On the other hand, governments who do not care about potential backlash from uninformed, broader audiences are not targeted with very many WTO disputes. The re-

[17] This weighting parameter is the α parameter from the well-known Grossman and Helpman (1994) model.

[18] This figure also excludes Singapore, which is a huge outlier in terms of its estimated $alpha$, 404.29. Results discussed below are substantively similar when including Singapore, the U.S. and E.U..

lationship presented in Figure 2.3 is statistically significant as well. In a Poisson regression of the number of disputes targeting a particular country on that country's estimated α, higher α is associated with being targeted by more WTO disputes, and is significant.[19] These results are consistent with existing analyses of the effect of democracy (measured by Polity scores) on the number of disputes experienced by a particular dyad (Sattler and Bernauer, 2011).[20]

What is the home government's optimal initial policy? The home government's initial optimization problem and related first order condition are:

$$max_{t_1} \quad \Pi(t_1)u_H(t_2^*) + (1 - \Pi(t_1))u_H(t_1)$$

$$max_{t_1} \quad F(t_1 - t_2^*)u_H(t_2^*) + (1 - F(t_1 - t_2^*))u_H(t_1)$$

$$[1 - F(t_1^* - t_2^*)]u_H'(t_1^*) = f(t_1^* - t_2^*)[u_H(t_1^*) - u_H(t_2^*)]$$

For a good home government, their optimal policy choice is $t_1^* = A$. Good home governments can do no better by choosing a different initial policy. If the foreign government draws a negative litigation cost and initiates a dispute, then the good home government will still choose $t_2^* = A$. If the foreign government draws a higher litigation cost, they will not initiate a dispute and the audience will not mobilize, leaving the home government's ideal policy in place.

Bad governments face a more complicated tradeoff. They can raise the initial tariff towards their ideal tariff level, which will be better for them if they avoid a dispute. But at the same time, choosing a higher initial tariff increases the probability of a dispute by increasing the relative attractiveness of a dispute to the foreign government. The first order condition shows how, in equilibrium, the marginal gain from raising the initial tariff, i.e. the marginal utility of the tariff

[19] The regression is a Poisson regression with the number of times a country has been a WTO respondent since 1995 as the dependent variable and the Gawande, Krishna and Olarreaga (2009) estimate of α as the independent variable, covering 38 countries for which the Gawande et. al. data are available. The coefficient (SE) are 0.05(0.006) and the associated p value is less than 0.00. These results are the same using a negative binomial regression or simple bivariate linear regression, as well.

[20] Other analyses have focused on the effect of regime type on the decision to *initiate* a dispute, which is distinct from the argument developed here, which describes the effect of government sensitivity on the likelihood of being *targeted* by a dispute.

times the probability of avoiding a dispute, equals the marginal costs, i.e. the additional probability of a dispute times the home government's utility lost from having to update its policy in the face of audience mobilization.

How do audience features affect the home government's initial policy choice? This question is particularly important because we are concerned with the affect of dispute settlement mechanisms on the tariff policies chosen both in the presence of disputes *and* when we do not observe disputes. If the presence of a dispute settlement causes governments to choose lower tariffs, even in the absence of disputes, then this supports the contention that dispute settlement mechanisms are an important component of how institutions affect member state behavior.

Proposition 4. *The home government's optimal initial policy, t_1^*, is: (i) increasing in A, (ii) decreasing in α, and (iii) increasing in H.*

Proof. First, observe that for bad governments, $t_1^* \in [t_2^*, B]$. The home government can do no better by choosing an initial policy higher than B: lowering the policy to B decreases the probability of a dispute and leaves them better off if they avoid a dispute. Similarly, the home government can do no better by choosing a policy lower than t_2^*: raising the policy to t_2^* lowers the probability of a dispute by decreasing the distance between t_1^* and t_2^* and leaves the home government better off if they avoid a dispute.

Rewriting the FOC for the maximization problem associated with t_1^* yields:

$$f(t_1^* - t_2^*)[u_H(t_2^*) - u_H(t_1^*)] + [1 - F(t_1^* - t_2^*)]u_H'(t_1^*) = 0$$

Since t_2^* is uninfluenced by t_1^*, we can rewrite the FOC as:

$$h(t_1^*)\frac{\partial t_1^*}{\partial t_2^*} + g(t_2^*) = 0$$

where $h(t_1^*)$ is the total derivative of the FOC with respect to t_1^* and $g(t_2^*$ is the total derivative of the FOC with respect to t_2^*. Rearranging yields:

$$\frac{\partial t_1^*}{\partial t_2^*} = \frac{-g(t_2^*)}{h(t_1^*)}$$

Substituting in the total derivatives, $h(t_1^*)$ and $g(t_2^*)$ yields:

$$\frac{\partial t_1^*}{\partial t_2^*} = \frac{f'(t_1^* - t_2^*)[u_H(t_2^*) - u_H(t_1^*)] - f(t_1^* - t_2^*)[u'_H(t_2^*) + u'_H(t_1^*)]}{f'(t_1^* - t_2^*)[u_H(t_2^*) - u_H(t_1^*)] - 2f(t_1^* - t_2^*)u'(t_1^*) + [1 - F(t_1^* - t_2^*)]u''_H(t_1^*)}$$

Since $f'(k) = 0$ for the uniform distribution, this equation can be signed by observing that $u'_H > 0$ and $u''_H < 0$ for all $t \in [A, B]$. Specifically, we now know that $\frac{\partial t_1^*}{\partial t_2^*} \geq 0$. This implies that t_1^* "inherits" the properties of t_2^* that are described in Corollary 1. □

Proposition 4 shows how audiences features can magnify or constrain the ability of dispute settlement mechanisms to affect member state behavior, *ex ante*. As the audience prefers lower tariff levels, the home government must make policy in the shadow of potentially more severe consequences from audience mobilization. The same is true for increasing or decreasing audience strength. Stronger potential audiences who prefer lower levels of tariffs make dispute settlement a stronger deterrent to higher initial tariffs for bad governments. In the domestic courts literature, this phenomenon has been referred to as "autolimitation" (Vanberg, 2005, 1998; Stone, 1992). When faced with the prospect of costly judicial review, legislatures may propose more moderate policies than they would have in the absence of any threat of judicial review. The same is true of governments facing the prospect of audience backlash resulting from a legalized international dispute. When audience punishment is more costly, governments choose more compliance policies *ex ante* in order to decrease the likelihood that they will face such punishment.

However, these results also show how the ability of dispute settlement to affect the home government's behavior is tempered by features of the audience. As the audience prefers higher tariff levels, the home government is less constrained by dispute settlement and chooses higher initial tariffs. Similarly, when facing weaker audiences, the specter of a dispute and potential audience mobilization is less frightening. This result shows that credible commitments and audience costs theories need to pay careful attention to the features of audiences receiving information. Theories that assume credibility and costliness result directly from the presence of the institution are incomplete at best. If the audiences who receive information from the institution do not support compliance or are too weak to impose costs even when they mobilize, then the effect of the institution on compliance behavior is weaker. Fundamentally, the institution cannot induce compliance levels that are higher, i.e. lower tariffs, than the relevant audience desires.

The effect of audience features on the home government's initial policy choice complicates a description of how audience features affect the equilibrium probability of a dispute, $\Pi(t_1^*)$. On the one hand, a more favorable audience from the Foreign government's perspective (audiences that are strong and like lower tariffs) makes a dispute *more* likely. Favorable audiences have a *post-dispute effect*, meaning that the Foreign government can induce larger changes in the Home government's policies, as was the intuition in Proposition 3. On the other hand, Proposition 4 says that more favorable audiences also have a *pre-dispute effect*. The Home government anticipates its audience's reaction when choosing its initial policy. Better audiences therefore lower the probability of a dispute by making the Home government choose lower tariffs initially.

Proposition 5 describes the conditions under which each effect dominates.

Proposition 5. *If* $f(t_1^* - t_2^*)u_H'(t_2^*) \leq -[1 - F(t_1^* - t_2^*)]u_H''(t_1^*)$ *then* $\frac{\partial \Pi(t_1^*)}{\partial A} \geq 0$ *and* $\frac{\partial \Pi(t_1^*)}{\partial \alpha} \leq 0$

Proof. This proof builds off of the proof for Proposition 4. Recall that the proof of Proposition 4 showed that $\frac{\partial t_1^*}{\partial t_2^*} \geq 0$. Now, we consider whether $\frac{\partial t_1^*}{\partial t_2^*} \leq 1$. If $\frac{\partial t_1^*}{\partial t_2^*} \leq 1$, then equilibrium increases in t_2^* result in *smaller* accompanying increases in t_1^*. Since k is distributed uniformly, this would imply that the post-dispute effect dominates.

Recall the expression for $\frac{\partial t_1^*}{\partial t_2^*}$ with the uniform distribution simplifies to:

$$\frac{\partial t_1^*}{\partial t_2^*} = \frac{f(t_1^* - t_2^*)[u_H'(t_2^*) + u_H'(t_1^*)]}{2f(t_1^* - t_2^*)u'(t_1^*) - [1 - F(t_1^* - t_2^*)]u_H''(t_1^*)}$$

Since the numerator and denominator have the same sign, for $\frac{\partial t_1^*}{\partial t_2^*} \leq 1$ it must be the case that:

$$f(t_1^* - t_2^*)[u'_H(t_2^*) + u'_H(t_1^*)] \leq 2f(t_1^* - t_2^*)u'(t_1^*) - [1 - F(t_1^* - t_2^*)]u''_H(t_1^*)$$

$$f(t_1^* - t_2^*)u'(t_2^*) \leq -[1 - F(t_1^* - t_2^*)u''(t_2^*)$$

□

Proposition 5 shows why careful attention needs to be paid to linking the occurrence of disputes with compliance. The effects of audience features on the equilibrium probability of a dispute depend on assumptions made about the curvature of the Home government's utility function, and implicitly about the shape of the litigation costs distribution function, as well. In some cases, stronger, compliance-favoring audiences increase the equilibrium probability of a dispute, and in other cases, they decrease this probability. An often-used dispute settlement mechanism may not be an effective one, if the frequency of its use is the result of its failure to deter initial violations. A rarely-used dispute settlement mechanism may, in reality, be the most effective. Governments refrain from violating their agreements too severely because they fear the possibility of a dispute.

One way to gain empirical leverage on the effects of audience features on the probability of a dispute is to consider how connected the pre- and post-dispute decisions are for the Home government. Empirically linking audience features to the probability of a dispute is most straightforward when the government's initial decision is distinct from its post-dispute compliance decision. In other words, if the pre-dispute effect of audience features is negligible, i.e. the Home government does not intensely anticipate possible audience reactions when making its initial decision, then we can apply the intuition of Proposition 3 to gain empirical traction.

There are many real-world situations that generate this type of separation between pre- and post- dispute decision-making. This separation occurs if different political actors make the pre- and post-dispute decisions. For example, in the context of U.S. antidumping and countervailing duty policy, private firms petition bureaucratic agencies like the Department of Commerce and International Trade Commission for tariff protection in the form of antidumping and countervailing duties (the pre-dispute decision). These duties have often been targeted as WTO-illegal in subsequent WTO disputes. Yet the handling of WTO disputes and any subsequent policy adjustments (the post-dispute decision) are handled by the Executive branch and the U.S. Trade Representative. To the extent that these bureaucratic agencies are not "perfect agents", the possibility of audience punishment does not affect the initial policy decisions of bureaucracies in the same way that it affects the executive's decisions.[21]

The length of time between many violations and subsequent disputes further disconnects the initial policy decision from the post-dispute policy decision. For example, a policymaker may erect a trade barrier even if they fear possible audience repercussions because they know that any dispute is likely to come much later, if at all. The policymaker may discount the audience's preferences in their initial decision, but be responsive to the audience after a dispute. Audience features also change after the government has chosen its initial policy. If audience features change *after* the initial policy decision, then government's might make policy according to the preferences and strength of their current audience, or expected future audience. But if those audience features changed in the future, that could make disputes more or less likely.

In these types of situations, where there is separation between the pre- and post-dispute decisions, the equilibrium probability of a dispute inherits the features described in Proposition 4. More favorable audiences make disputes more likely. With the initial violation already committed, foreign governments observe audience features and decide whether the audience is "ripe" for activation with a dispute. Empirically, this appears to be the case. Chaudoin (2011) shows how the timing of trade disputes against the United States is consistent with this theory. The United State's trade partners are more likely to initiate WTO disputes during low-unemployment election years. In other words, they litigate against the U.S. when the audience is more amenable to free trade (lower unemployment, better macroeconomic conditions) and when policymakers actually care about these broad constituencies (during elections).

[21] American politics literature has a rich history of studying delegation to bureaucratic agencies and the degree to which agents behave according to the wishes of their principals (McCubbins, Noll and Weingast, 1987).

Further Implications and Conclusions

This paper delivers good news and bad news for theories of international institutions that are based on noncompliance costs, like theories of credible commitments or audience costs. The good news is that institutions can generate these types of costs vis-a-vis domestic audiences under very minimal restrictions. The institution need only provide a costly way for a foreign government to signal to domestic audiences in a home country that the home government has misbehaved. Dispute settlement bodies provide such a forum since their use is both costly and public. When the preferences of the foreign government and the home audience are at least partially aligned, such a mechanism can help the home audience better constrain their government and deter the home government from choosing policies that are at odds with its international obligations, even when disputes do not occur.

This is especially good news since related theories in the context of security and conflict, namely audience costs theories of crisis bargaining, have recently taken a beating. On the theoretical side, Slantchev (2006) argues that audience costs exist in crisis bargaining situations only under certain restrictive conditions. On the empirical side, Snyder and Borghard (2011) find very little empirical evidence for audience costs and question its assumptions in a variety of ways. Noncompliance costs in the institutions context are closely related to audience costs in the crisis bargaining context. Both are costs (potentially) incurred by a government for breaking its promises, either the implied promise of military action when a leader makes a threat or the more explicit promises codified by international agreements.

How is it that these theories appear coherent in the institutions context, when they have received such criticism in the crisis bargaining context? The answer lies in a key differences between the contexts. In crisis bargaining, whether a government follows through on its promises, i.e. takes military action when its opponent does not capitulate to threats, is a public act. In the context of international agreements, whether or not a government complies with its obligations is far from public. Compliance behavior in virtually every context governed by international agreements (trade, human rights practices, environmental protection) is very difficult to monitor, even if a particular audience was willing and able to punish its government for noncompliance. This creates the unique potential for dispute settlement mechanisms to act as costly signaling devices that can play a key role in alleviating the informational disadvantage suffered by audiences.

However, this paper also delivered bad news for theories of noncompliance costs in the context of international institutions. The credibility of credible commitments and the costliness of audience costs are far from guaranteed. Specifically, the ability of the institution to generate these costs is constrained by the preferences and political strength of the audience in question. The institution cannot take compliance further than the audience is willing to go. At one extreme, when the audience supports non-compliance, providing them information about their government's decision can potentially create incentives to decrease compliance. Less extreme, though still troubling, is the fact that audiences who only weakly prefer compliance or who are politically weak do not generate significant noncompliance costs, and therefore do not constrain their government from misbehaving.

This paper also generated empirically testable predictions about pre- and post-dispute behavior of governments, as well as the likelihood of observing a dispute. Pre- and post-dispute government behavior reflect features of the audience. Audiences that strongly prefer more compliant policies place a tighter leash on their government, and as a result, make their government more compliant with its obligations. The likelihood of observing a dispute depends on the degree to which governments anticipate audience punishment in their initial noncompliance decisions.

This analysis also explained a puzzling empirical phenomenon: if democratic governments are thought to be more likely to honor their international agreements (i.e. with lower tariffs, (Milner and Kubota, 2005)), then why are they so much more likely to find themselves as defendants in front of international dispute settlement bodies? This phenomenon arises, at least in part, because democracies are the ones for whom the information-transmission role of dispute settlement is most effective. Autocracies, whose audiences are less able to mobilize and influence their government's policies, even when they have sufficient information about their government's behavior, do not get litigated against because transmitting information to their audiences is less likely to justify the costs of litigation for a potential plaintiff.

Finally, this paper weighs in directly on a prominent debate over whether multilateral organizations are democracy-enhancing or hindering (Keohane, Macedo and Moravcsik, 2009, 2011; Gartzke and Naoi, 2011). As Keohane, Macedo and Moravcsik (2009) highlight, the debate over multilateral organizations (MLO's) is often between those touting their pragmatic benefits and those critical of the possibility that MLO's can undermine democracy. Keohane, Macedo and Moravcsik (2009) describe a variety of mechanisms through which MLO's can actually enhance democracy. A key battlefield in this argument, and in subsequent criticism (Gartzke and Naoi, 2011) and response (Keohane, Macedo and Moravcsik, 2011), is over the concept of representation. This paper demonstrates a potentially powerful way that institutions can enhance representation, via their ability to provide information. Government policymaking is often an exercise in balancing competing special interest groups, with the preferences of broader audiences often receiving less emphasis than the preferences of concentrated, well-organized groups. In part, this stems from the informational advantage held by special interest groups. They can better monitor government policy and better choose when to mobilize against policies that are contrary to their interests and when to husband their resources when their government chooses policies in line with their interests. Broad, diffuse audience do not have this luxury. However, if international institutions, and features like public and costly dispute settlement bodies, can lessen the informational advantage of special interest groups, then this is a clear improvement in the representativeness of policies chosen by elected officials.

Figure 2.1: Effect of Audience Preferences on Optimal Post-dispute Policy

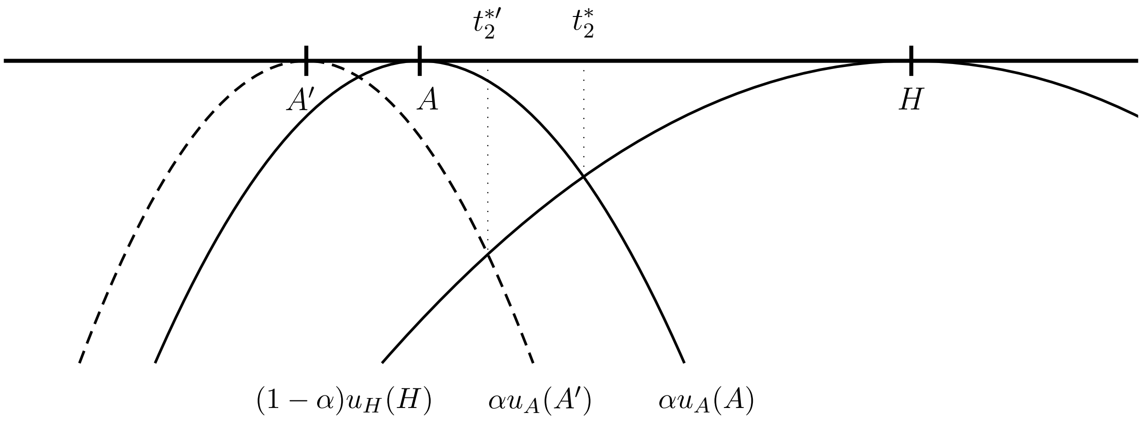


Figure 2.2: Effect of Audience Strength on Optimal Post-dispute Policy

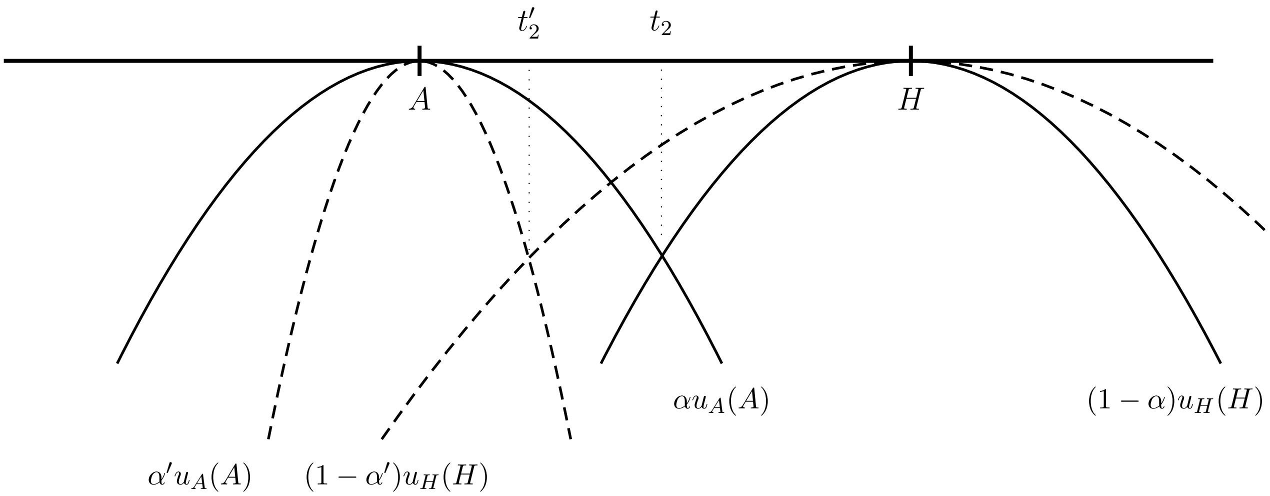
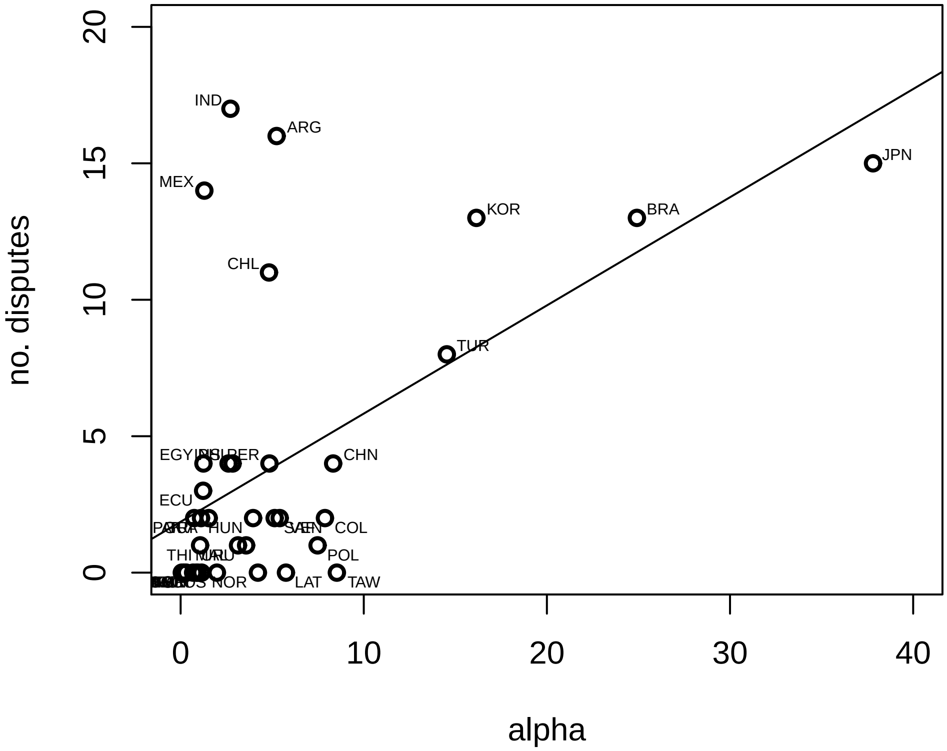


Figure 2.3: No. WTO Disputes vs. Estimated Alpha



Horiz. axis is the estimate of alpha from Gawande et. al. (2009). Higher values indicate greater concern for aggregate welfare. Vert. axis is the number of WTO disputes against the country. For fitted regression line, Coeff. = 0.40, SE = 0.05. Excludes the U.S. (alpha = 26.14, 95 disputes), E.U. (av. alpha = 9.49, 67 disputes) and Singapore (alpha = 404.29, 0 disputes).

Chapter 3

Macro-level Evidence: The Strategic Timing of Trade Disputes

Legalized dispute settlement mechanisms (DSMs), which increasingly accompany international agreements, are thought to play a particularly important role in facilitating international cooperation because of their ability to transmit information about the behavior of member states (Mansfield, Milner and Rosendorff, 2000, 2002; Rosendorff, 2005). When one member state violates an agreement, another member state can use the agreement's DSM to sound the alarm. Hearing the alarm, domestic audiences punish the offending government. The alarm mechanism raises the costs of defection, which makes cooperation more attractive *ex ante*. While not always linked to dispute settlement, the ability of international institutions to raise the costs of defection by informing, activating, or mobilizing subnational actors is at the core of existing theories of international cooperation based on credible commitments (Simmons, 2000; Simmons and Danner, 2010) and audience costs (Tomz, 2007).

Despite the benefits of sounding the alarm, the occurrence of disputes is neither immediate nor automatic. At virtually every DSM forum, there is significant variation in when disputes occur. Often, a large amount of time elapses between when one member state violates an agreement and when the aggrieved member state initiates a legalized dispute, if any dispute occurs at all. If institutions, and DSM's in particular, are important alarm mechanisms, then why do member states wait months, or even years, before sounding the alarm?

The World Trade Organization's (WTO) Dispute Settlement Understanding (DSU) is one of the world's most sophisticated and important forums for settling interstate disputes over trade barriers. Under the DSU, members regularly request consultations and bring formal litigation against one another over real (or perceived) violations of WTO law. To date, over four hundred cases have been brought before the WTO's DSU, covering diverse issue areas of international trade. Yet, even in such an advanced DSM, there is significant variation in whether and when a dispute occurs. When a WTO member is suspected of violating WTO law, disputes sometimes occur quickly, within months of the perceived violation. Other times, one country will wait years before ever challenging a particular trade barrier enacted by another country. Many WTO-inconsistent policies never receive any scrutiny at the DSU. Even apart from broader theories of institutions and cooperation, this is important variation in search of an explanation because these disputes affect significant international trade flows and the disputes themselves impose significant costs on the litigants.

I argue that features of the audience who gains information from a dispute influence the costs and benefits of a dispute, and subsequently dispute timing. For the plaintiff country, a dispute is valuable largely because of the prospect of changing the defendant's policies. The reaction of audiences to the dispute potentially compel the government to bring its policies in-line with the desires of the plaintiff and the terms of the international agreement. The audience's preferences and political strength affect the magnitude and direction of that reaction. From the plaintiff's perspective, a dispute has the best chance of changing defendant government practices when the listening audience is strong and supports compliance. If a dispute informs the audience that their government has chosen policies contrary to audience preferences, then their reaction might be to punish their government with the goal of changing its policies. If the dispute informs the audience that their government has chosen policies in-line with the audience's preferences, then the audience is unlikely to react negatively. The effect of audience reaction is magnified by the political strength of the audience; the reactions of strong audiences are more important than the reactions of weak audiences. Disputes, therefore, should be most likely when audiences are "willing and able" to encourage the defendant government to comply.

I test this prediction by modeling the timing of disputes against a particularly important subset of U.S. trade policies, antidumping (AD) petitions and countervailing duties (CVD's). AD and CVD petitions have been the targets of a large portion of the WTO's caseload (Bown, 2004) yet are relatively opaque policies. I find that the timing of WTO challenges to AD and CVD tariffs is consistent with the above predictions. U.S. tariffs are more likely to be targeted by WTO disputes when broader U.S. audiences are willing and able to support free trade policies. As national elections approach, disputes are more likely when macroeconomic indicators, like unemployment rates, suggest support for free trade. U.S. trading partners tend to delay disputes as elections approach when unemployment is high.

This paper is the first (to my knowledge) to empirically analyze the *timing* of disputes as opposed to their number or occurrence (Bown, 2005*b*; Davis, 2011; Davis and Shirato, 2007; Davis and Bermeo, 2009; Sattler and Bernauer, 2011). Given the significant variation in the timing of disputes, understand *when* they occur is as important as understanding *whether* they occur. The results show that dispute timing is consistent with predictions derived from alarm theories of DSM's. The alarm mechanism is an important role for institutions, but its effect of member state behavior is constrained by features of the audience hearing the alarm. I also show that functionalist explanations for the effects of international institutions on member state behavior, like those based on information transmission, can operate in simultaneously with more realist explanations based on power politics. When explaining whether or not "institutions matter," the answer need not be a stark yes or no. Both institutional effects and power politics can be important explanations for member state behavior.

Information and Domestic Audiences

One theoretical explanation for how international institutions affect member state behavior is that institutions transmit information to particular audiences, who can punish their leaders for defections from international agreements. In audience costs theories, the audience is often thought of as a set of voters and punishment is electoral: leaders who break their international agreements are not returned to office (Tomz, 2007; Mansfield, Milner and Rosendorff, 2000). In her theory of credible commitments, Beth Simmons (2000) argues that relevant audiences consist of private economic actors. A government's IMF Article VIII obligation "mobilizes a new set of external actors (private economic, governmental, and legal) who may exert pressure to comply on a government that is considering or engaging in rule violation" (pg. 821). These types of arguments have been made in a variety of contexts, ranging from international trade agreements (Bthe and Milner, 2008; Mansfield, Milner and Rosendorff, 2000) to bilateral investment treaties (Elkins, Guzman and Simmons, 2006) to human rights (Simmons, 2009) to war crimes (Simmons and Danner, 2010).

The informational problems facing domestic audiences in the context of trade policy are particularly acute. Audiences may know whether their government supports free trade broadly, but likely cannot monitor opaque trade policies like antidumping petitions, countervailing duties, or non-tariff barriers. Alexandra Guisinger (2009) argues that voters in congressional elections often did a poor job of matching their preferences with their voting patterns in regards to the Central American Free Trade Agreement. Daniel Kono (2006) argues that democracies deliberately choose "optimal obfuscation" for their trade policies, in order to avoid electoral punishment. In the United States, relatively obscure trade policies like antidumping petitions get very little media attention, even though they are important trade policy tools. A search of U.S. newspapers from January 1999 to January 2012 for the terms "'antidumping' within 100 words of 'United States," yields only 390 results.[1]

Dispute settlement can play a key role in transmitting information to subnational actors and ameliorating this informational problem (Dai, 2002, 2007; Fang, 2008; Mansfield, Milner and Rosendorff, 2002). When a government violates the treaty, another member state can initiate a dispute, which "sounds the alarm" that a violation has occurred. Equipped with this new information about the behavior of their government, audiences can punish or reward their elected officials accordingly.

The WTO's DSU acts as just such a fire alarm and information clearinghouse (Rosendorff, 2005). One member state government, the "complainant," can formally request consultations over objectionable practices of another member state, the "respondent." The disputants attempt to negotiate a solution, but if they fail to reach a resolution, they can request the establishment of a Dispute Resolution Panel, which hears the case and issues a ruling. If the respondent loses and does not bring their trade policy in line with the panel's ruling, the panel approves compensation for the complainant, usually allowing them to raise their own trade barriers against the respondent.[2]

Two features of the WTO's DSU process make it particularly important for information transmission: the ability of the WTO to heighten awareness about violations and the costliness of litigation. In discussing a 1999 dispute between Chile and Argentina over vegetable oil tariffs, Tussie and Delich (2005) write: "Although accessible only to highly profitable sectors because participation is too costly and time consuming, the WTO provides the intangible benefit of *exposure*. Pressure through exposure can help countries unable or unwilling to retaliate to obtain more favourable results than in bilateral or regional instances. In fact, WTO rulings act as a *magnifying glass* of countries' (WTO-incompatible) trade policies" (23, emphasis added). Tellingly, the two parties initially sought to address this dispute under the auspices of MERCOSUR, yet these efforts failed. While they do not explicitly attribute the failure of the MERCOSUR process to these rea-

[1] Search conducted in Lexis Nexis Academic using "U.S. Newspapers and Wires" and the terms "antidumping! w/100 united! state!" on March 20, 2012.

[2] This is a slight simplification. Parties can request an appeal, and the Appellate Body can evaluate the actions of the original panel.

sons, Tussie and Delich (2005) note that the MERCOSUR efforts were "out of the public eye and at the same time it was both fast and low-cost" (30).

Litigation costs are another important feature of dispute settlement because they force potential plaintiffs to be strategic. Because DSU disputes consume immense amounts of resources, member states cannot simply initiate a dispute over every instance of protectionism by another member state.[3] Significant litigation costs associated with WTO disputes make them a costly and more credible signal.[4] Since disputes are so costly, no government can afford to "cry wolf" every time they want to accuse another member state of violating WTO rules.

The DSU's role in heightening awareness was evident in international efforts to address one particularly opaque, yet important, U.S. trade policy concerning the practice of "zeroing." Zeroing refers to the accounting procedures used when U.S. bureaucracies calculate whether to impose tariffs on certain imports and how large those tariffs should be.[5] Zeroing had long been a source of contention between the U.S. and many of its trading partners, and it been used since long before any other countries challenged its legality at the WTO.

Zeroing came to play an important role in several high profile WTO disputes with the European Communities and others.[6] Until other countries decided to object to zeroing using a formal WTO dispute, however, media coverage of this issue was virtually non-existent. Media coverage of

[3] By one estimate, a typical WTO dispute costs the litigants one million dollars apiece, which is a nontrivial sum when considering the size of the bureaucracies charged with handling WTO litigation, especially in small countries. Litigating disputes also takes time, which entails an opportunity cost of using litigation resources for other potential violations Davis and Shirato (2007). For countries unfamiliar with the DSU process, gaining experience about this legal arena entails the start-up costs of learning to argue effectively in front of the DSU Davis and Bermeo (2009).

[4] This is similar to the arguments made by Christina Davis (2011) in discussing how governments often engage in WTO disputes as a costly way for the government to reassure domestic firms that the government is committed to defending the firms' interests. Counter-intuitively, the fact that WTO disputes are expensive, and thus a more costly signal, may make them more credible information transmission mechanisms than other options. For instance, if an aggrieved country simply issued a press release bemoaning another country's trade policies, no one would pay attention because this action would be costless, and there are clear incentives to misrepresent.

[5] When U.S. bureaucracies investigate whether or not another country has sold goods on the U.S. market at below market price (i.e. dumping), they calculate dumping margins, or the amount below fair market price that the goods are being sold, across different companies and countries. When a particular firm is actually selling the goods at above market price, which would result in a negative dumping margin, the U.S. "zeroed" these margins, rounding them down to zero, and artificially inflating the amount of dumping that was occurring. For a more extensive review see: Alford (2006).

[6] Argentina, Brazil, China, Taiwan, Hong Kong, India, Japan, South Korea, Mexico, Norway, Turkey and Canada were also involved in disputes with the United States over zeroing in some fashion, either as third parties or by

zeroing did not begin until June of 2003, shortly after the European Communities initiate legal WTO action against the United States over the practice. After that, media coverage of zeroing increases sharply, with coverage even reaching the pages of such well known publications as the New York Times and Washington Post.[7]

While zeroing was a particular trade issue that attracted international attention, the same pattern holds when looking at other trade policies. The number of hits for the Lexis Nexis search described above decreases from 390 to 265 when the term "World Trade Organization" is excluded. About a third of the coverage of antidumping petitions includes some reference to the WTO. Chang, Golden and Hill (2010) argue that increased media coverage of the behavior of politicians goes a long way towards helping the electorate hold politicians accountable. The additional exposure resulting from WTO disputes raises the profile of previously opaque trade policies.

The Puzzle

If dispute settlement is an important way that international institutions can transmit information and activate domestic audiences, then a crucial question is: when should these disputes occur? If disputes sound an alarm that causes the defendant to change their policies, then plaintiffs should initiate disputes quickly.[8] Delay only lengthens the amount of time that the plaintiff suffers from the defendant's non-compliant policies. Yet variation in timing of disputes is the norm, rather than the exception. Plaintiffs rarely initiate disputes immediately after violations occur, and the amount of time between the defendant's alleged violation and the plaintiff's resulting decision to initiate a dispute is often significant.

challenging the practice of zeroing in other DSU disputes. Zeroing became an important issue especially in Canadian complaints against U.S. tariffs on imports of Canadian lumber.

[7] The initial search used the terms "united states and dumping and zeroin! and commerce" in Lexis Nexis Academic Universe, in US Newspapers and Wires and Major Newspapers, searched on 10/05/10. The two articles referenced are "A Trade Battle is Brewing Over U.S. Antidumping Fees," *New York Times* 2/18/2004 and "Jumbo Shrimp Follies," *The Washington Post* 11/15/2004. There are over 100 hits using those search terms that occur after June 2003. The first mention of zeroing is in "European Commission Protests US Method Of Calculating Anti-Dumping Fees," *The White House Bulletin* 6/13/2003.

[8] For clarity and consistency, I will refer to parties as the "plaintiff" and "defendant" even though they may have different names under different DSM's.

To see this variation in timing of disputes, consider the United States' practices regarding AD and CVD petitions. In the United States, domestic producers can file petitions with particular federal bureaucracies, the International Trade Commission (ITC) and Department of Commerce (DOC), when they suspect that exporters from foreign countries are "dumping:" selling products in the United States at below market price either because of predatory pricing or subsidization by the foreign government. After a U.S. firm files a petition, the relevant bureaucracies evaluate whether dumping is indeed occurring and whether the U.S. firm has been harmed as a result. If so, they issue an affirmative preliminary ruling, and places tariffs on the goods in question.[9] The bureaucracies and petitioning firm then enter into a lengthier evidence-gathering phase in order to make a final ruling. If the bureaucracies issue affirmative final rulings, the preliminary duties stay in place until they expire or are revoked when dumping is deemed to have ceased. Petitions are very successful at the preliminary stage, with the majority receiving an affirmative preliminary ruling.

The tariffs resulting from AD and CVD petitions have been a particularly contentious issue at the DSU. Disputes concerning AD and CVD petitions make up a large part of the DSU's caseload, and in virtually every case concerning these tariffs, the WTO has ruled in favor of the complainant on at least one issue in the case (Bown 2005, 516-517). AD and CVD cases also account for a large proportion of the WTO litigation targeting the United States: of the 111 instances in which the United States has been named as a respondent in a WTO dispute since 1995, 42 (approx. 38%) were focused primarily on AD and CVD actions.[10] The AD and CVD processes have thus often generated DSU-actionable trade barriers and foreign governments largely have been successful in their legal challenges.

Yet there is significant variation in the timing of DSU disputes against AD and CVD tariffs. Figure 3.1 shows the distribution of the length of time elapsing between when an AD or CVD peti-

[9] The CVD process is slightly different from the AD process, but they are similar enough for the analysis here. The description here most closely describes the AD process.

[10] This tally actually understates the importance of AD and CVD petitions to the United States' experience with the DSU since I only counted disputes which specifically referenced AD or CVD in their official WTO DSU title.

tion received an affirmative preliminary ruling and the foreign government harmed by the resulting tariffs chooses to initiate a DSU dispute over that tariff.[11] Some tariffs are challenged relatively quickly; the foreign government requests DSU consultations within a few months of the affirmative ruling. Other tariffs are in place for years before the foreign government challenges them at the DSU.

Why would the foreign government whose exporters were harmed by AD and CVD tariffs wait before challenging them at the DSU? The tariffs directly harm the interest of foreign exporting firms, and the petitions can have significant chilling effects on a country's aggregate imports (Vandenbussche and Zanardi, 2010).

Existing explanations for DSU disputes focus on explaining the *occurrence* of disputes rather than the *timing* of disputes. For example, Sattler and Bernauer (2011) argue for a "gravitational explanation:" dyads involving larger countries with larger trade flows between them experience more DSU disputes. Yet the sizes and trade flows among dyads is relatively constant over time. By and large, big countries stay big, and small countries stay small. The trade intensity of dyads relative to other dyads is also fairly constant. Even within dispute-prone dyads, like the United States-European Communities dyad, there is significant variation in the timing of DSU disputes that cannot be explained by the countries' sizes or trading intensity.

Legal explanations are also important for dispute occurrence. For example, a country's legal capacity may affect its ability to initiate disputes (Busch and Reinhardt, 2003; Busch, Reinhardt and Shaffer, 2009; Guzman and Simmons, 2005; Horn, Mavroidis and Nordstrom, 1999). High-capacity countries initiate more disputes. Yet, legal capacity is also fairly time-invariant. There is rarely significant variation in a country's legal capacity from month to month. Other legal explanations that focus on forum shopping (Busch, 2007) are also ill-equipped to explain the timing of disputes since the relative attractiveness of different venues does not vary significantly over time.

[11] This figure is limited to the petitions that received affirmative rulings after April 1994 and were petitions against WTO members, since only WTO members can use the DSU.

Power politics are also important. Countries use trade retaliation to punish other countries' noncompliance with WTO-rulings. Plaintiffs should be most likely to initiate DSU disputes when the defendant is most sensitive to direct punishment by the plaintiff and when the plaintiff is most immune from "counter-retaliation." The threat of this punishment is more potent when the defendant exports more to the plaintiff. If the defendant exports nothing to the plaintiff then the plaintiff has no defendant-exported goods to "hold hostage." Retaliation after a dispute also risks counter-retaliation cycles. The threat of a retaliatory trade war from the defendant is more acute when the plaintiff exports more to the defendant. Conversely, if the plaintiff does not rely at all on exports to the defendant, then the respondent has no "counter-threat" with which to deter litigation. Bown (2005*a*) finds that retaliatory capacity is an important determinant for participation in WTO litigation, even when controlling for other important quantities like the amount of exports at stake.

Finally, lobbying by firms in the plaintiff country is also important (Davis and Shirato, 2007; Davis, 2011). Some countries file disputes in order to placate domestic firms who have been harmed by foreign barriers. These explanations emphasize across-firm variation: firms in "static" industries who can tolerate the lengthy DSU process more strongly lobby their government for litigation. However, these cross-firm or cross-industry characteristics are also largely time-invariant.

The Argument

Features of audiences in the defendant country are important determinants of the timing of disputes. Specifically, the preferences and political strength of those audiences affect how the audience reacts to a dispute, and in turn, whether or not a dispute is an attractive option for the complainant. When a plaintiff can use a dispute to mobilize, activate, or inform a strong, compliance-supporting audience in the defendant country, then the benefits of a dispute are more likely to outweigh the costs. When a dispute could potentially mobilize a hostile audience or would only succeed in mobilizing a weak audience, the expected value of a dispute drops, from the plaintiff's perspective.

Consider variation in the preferences of domestic audiences. The WTO proscribes free trade; its “overriding objective is to help trade flow smoothly, freely, fairly and predictably.”[12] Yet the general public rarely, if ever, supports zero protectionism. The degree of support for free trade often varies with macro-economic conditions. Bergsten and Cline (1983) describe how “high levels of unemployment are the single most important source of protectionist pressure.” Mansfield and Busch (1995) find that higher unemployment is associated with increased non-tariff barriers since unemployment creates demands for protection. Mansfield and Mutz (2009) argue that perceptions of trade policy’s effects on the economy as a whole affect individuals’ attitudes about free trade. Support for free trade may wax and wane over time, especially as macroeconomic conditions improve or worsen.

Also consider variation in how much the defendant government cares about possible audience reactions. Governments vary over time in the degree to which they are sensitive to the preferences of broader audiences. One motivation for the vast amount of literature on political business cycles is that argument that during an election year, politicians are more sensitive to the general public (Nordhaus, 1975). As elections approach, politicians choose policies with the goal of reaping electoral reward. When the specter of elections does not loom as large, politicians are more free to pick policies that diverge from the preferences of their electorate.

Existing research has focused on cross-national variation in leaders’ sensitivity to public preferences. James Fearon (1994) original argument about audience costs emphasizes cross-national variation in regime type. Democracies are more susceptible to audience punishment than non-democratic regimes, and this affects crisis bargaining behavior. Allee and Huth (2006) find that democracies are more likely to use legal dispute settlement for territorial disagreements.

Combining variation in audience preferences and strength yields a conditional hypothesis: when the defendant government is sensitive to audience preferences *and* when those audiences prefer free trade, disputes should be more likely. When the defendant government is sensitive to

[12] Quoted previously in Rose (2004), from $http://www.wto.org/english/rese/doloade/inbre.pdf$

audience preferences and those audiences are more supportive of protectionism, disputes should be less likely. Disputes should be more likely to delay disputes until the defendant government is less sensitive to pro-protectionism audiences.

Modeling the Timing of Disputes

To test whether audience features affect the timing of DSU disputes, I use data describing the lifespan of trade barriers resulting from United States AD and CVD petitions. The key feature of the data is that they track a set of potential DSU disputes and describe whether *and when* those disputes occurred. I first use Chad Bown's "Global Antidumping Database" and extract all of the AD and CVD the petitions filed by U.S. firms from April of 1994 to October of 2009. Each observation in the Bown dataset describes one particular petition and contains information on the time of its initiation, the target country, the products affected, the rulings of the relevant U.S. bureaucratic bodies at the various stages of the process, the dates of these rulings, and any resulting WTO litigation.[13] The choice of the starting date reflects important institutional changes to the WTO. April of 1994 marks the date of agreement for the transition from the old GATT regime to the new WTO regime, which included significant changes designed to strengthen the dispute settlement mechanism. These changes went into effect in January of 1995. I exclude AD/CVD petitions filed earlier in order to hold the institutional rules of the dispute settlement mechanism fixed throughout the analysis. I also excluded petitions that were filed against countries that were not WTO members at the time of filing. This ensures that the foreign country targeted by the petition is able to initiate a DSU dispute against the United States for the entire lifespan of the petition.

I break each AD/CVD petition down into monthly observations, so the unit of observation is the petition-month. I first begin observing a petition in the month that it receives the necessary

[13] This is just part of the information contained in this extensive dataset. It covers many other countries as well as other trade policies like safeguard actions. Its scope, comprehensiveness, and public availability are impressive and appreciated. The website for this data is $http://people.brandeis.edu/\ cbown/global\_ad/$.

affirmative preliminary rulings, and is awaiting a final ruling. As described above, this is the first stage of a petition's lifespan in which tariffs are applied. Petitions that do not pass the necessary preliminary rulings do not result in tariffs.[14] For clarity, I refer to petitions that have received affirmative preliminary rulings as tariffs.

After a petition receives an affirmative preliminary ruling, the resulting tariff can experience one of three possible events over the course of its lifespan: a WTO dispute, a negative final ruling, or revocation. A *WTO Dispute* occurs in the month in which the country targeted by a particular AD/CVD tariff formally requests DSU consultations over that tariff. A tariff can also receive a negative final ruling from the relevant U.S. bureaucracies, which terminates the tariff. Some petitions receive affirmative preliminary rulings only to receive negative final rulings after the evidence-gathering stages. A tariff can also be revoked if the relevant bureaucracies determine that tariffs are no longer warranted.

I group the final two events, negative final ruling and revocation, together and label them as *Unilateral Removal*, because these events both stem from decisions made by U.S. actors. A *WTO Dispute*, on the other hand, is a decision made by foreign actors. I draw the distinction between *WTO Dispute* and *Unilateral Removal* because it allows me to examine whether the effects of the explanatory variables differ across the type of event under consideration. *WTO Dispute* and *Unilateral Removal* are called "terminating events:" and I do not observe tariffs after either terminating event has occurred.[15] If neither terminating event occurs in a particular month, the tariff is labeled as *In Effect*, and it is possible for a tariff to still be in effect at the end of my observation time period, October of 2009.

[14] For the petitions that received affirmative preliminary rulings before January of 1995, I only begin observing these petitions in January of 1995, since this is when aforementioned institutional DSU changes go into effect.

[15] In practice, petitions can also be withdrawn by the petitioner. In these data, the only instances of withdrawal of petitions against WTO members occurred before preliminary rulings, which is before I begin observing the petition, so I do not consider this as a separate event. For a more extensive analysis of withdrawals, see: Prusa (1992).

The dependent variable, Y_{it}, therefore, is a categorical variable describing the "status" of the tariff i in month t. Y_{it} takes on a distinct numerical coding depending on whether the tariff is *In Effect* or experiences a *WTO Dispute* or *Unilateral Removal*.[16]

The 574 tariffs combine for 36,697 total months of observation time. Of the 574 tariffs, approximately 14% (78 tariffs), resulted in a WTO dispute before October of 2009. Approximately 55% (318 tariffs) ended because of unilateral bureaucratic decisions. Tariffs that resulted in a WTO dispute were in effect for approximately 77 months, with a minimum of 8 and a maximum of 252. Tariffs that were removed unilaterally were in effect for an average of 96 months, with a minimum of 10 and a maximum of 294.

Main Explanatory Variables

The theory's main prediction is that disputes are more likely when domestic audiences support free trade and when the U.S. government is most sensitive to those preferences. To proxy for domestic support for free trade, I use the U.S. unemployment rate. As described above, unemployment is the one of "usual macroeconomic suspects" associated with general support for free trade. *U.S. Unemployment* is a six month moving average of the monthly, seasonally adjusted percentage unemployed for people age 16 and over in the United States.[17]

To proxy for the government's sensitivity to support for free trade, *U.S. Election Year* is an indicator variable that is coded 1 in the twelve months preceding the next U.S. Presidential election, and zero otherwise. I focus on Presidential elections because the bureaucracies involved in AD and CVD petitions are most closely tied to the executive branch. Additionally, executives are thought

[16] In the parlance of survival models, each tariff is a particular subject. A subject is "born" in the month when the petition passes its preliminary rulings and is awaiting a final ruling. A subject "dies" or fails in the month that it experiences a terminating event. Subjects that do not experience any terminating events before the end of the observation window can be thought of as right-censored. Petitions filed before January 1995 but after April 1994 are left-censored until January 1995.

[17] Unemployment data are from the Bureau of Labor Statistics website, $http://www.bls.gov/$, Series ID: LNS14000000, and were accessed on February 16, 2010. The moving average includes the current month and the five preceding months. I use moving averages to capture broader economic trends, rather than transitory shocks. Results do not change if I use one month or twelve month moving averages for all the variables that are averaged.

to be responsive to broader constituencies than more narrowly-interest legislative members. Since the theory makes a conditional prediction for these two variables, I interact *U.S. Unemployment* and *U.S. Election Year*. During election years, higher unemployment should be associated with a lower probability of a WTO dispute.

The theory does not make predictions about the effects of unemployment and elections on the probability of *Unilateral Removal*. As explained shortly, this is an attractive feature of my approach, since it creates an informal and useful placebo test of the theory described above. If unemployment and elections have the predicted effect on the probability of a WTO dispute, but do not have the same effect on the probability of unilateral removal, then the results are more supportive of the theory.

Alternative Explanatory Variables

I also include variables to test for alternative explanations for the occurrence of trade disputes. The first two variables measure the potential for retaliation- where country A raises tariffs against country B's exports as punishment for B's tariffs. As described above, if the defendant exports a large amount to the plaintiff, disputes should be more likely since the plaintiff has greater leverage. When the plaintiff exports more to the defendant, they have less leverage. Retaliation should also increase the probability of unilateral removal. Blonigen and Prusa (2001) show that the possibility of retaliation decreases the probability that U.S. bureaucracies rule in favor of firms seeking protection. *U.S. Exports* measures the percentage of U.S. exports that go to the foreign country and *U.S. Imports* measures the percentage of U.S. imports that come from the foreign country.[18].

The second set of alternative explanations account for plaintiff-side dynamics. I include the most commonly used proxy for a country's legal capacity: their per capita GDP. The data for *Plaintiff PCGDP* come from the World Development Indicators dataset, measured yearly.[19] Macroeco-

[18] Again, I use six month moving averages. Trade data are from the U.S. International Trade Commission, $http$: $//dataweb.usitc.gov/scripts/INTRO.asp$.

[19] Busch, Reinhardt and Shaffer (2009) use survey data to construct a detailed measurement of legal capacity, but it is only cross-national and not time-series.

nomic and electoral dynamics in the plaintiff country may also affect the probability of a dispute. *Plaintiff Election* is an indicator variable that is coded 1 if the foreign country is within 12 months of its next major election, and zero otherwise. *Plaintiff Unemployment* codes the unemployment rate for the plaintiff country. As with the predictions for U.S. elections and unemployment, I also include their interaction.

Note, An important feature of the variables described above is that they are time-varying. Bown (2005*b*) and Horn, Mavroidis and Nordstrom (1999) argue that the stakes of the case are important. If a country can expect to regain a larger amount of its exports should the offending tariff be lifted, then they are more likely to initiate a dispute. The legal strength of a case also affects whether or not to file a dispute. If a country does not expect to win, then they will be less likely to file a dispute. I do not explicitly test these arguments here because they are largely time-invariant explanations. This is not to say that they are unimportant, but rather, they are less likely to explain variation in the *timing* of disputes.

Empirical Models

I estimate the effects of the explanatory variables on the status of a tariff (*In Effect*, *WTO Dispute*, *Unilateral Removal*) in two ways. First, I use a Cox proportional hazards model to estimate the effect of the variables on the risk of *WTO Dispute*. This approach is often used in modeling time-until-failure data. Specifically, I estimate the risk of a *WTO Dispute* for tariff i at time t: $h(t|X_{it}) = h(t)exp(X_{it}\beta)$.[20] This approach has the advantage of being able to estimate the effects of the explanatory variables on the risk of a *WTO Dispute*, while leaving the underlying, or baseline risk, of a *WTO Dispute* during time t, $h(t)$ unspecified.[21]

[20] The dependent variable is an indicator variable that equals 1 if the tariff experienced a *WTO Dispute* during that month, and zero otherwise.

[21] Note that time, t, is measured from the month that the petition receives an affirmative preliminary ruling, i.e. $t = 1$ refers to the first month of a tariff's lifespan. This is distinct from calendar time. I will control for possible trends in calendar time by including cubic polynomials that measure calendar time, i.e. *Month* $= 1$ refers to the first month in the sample (January of 1995).

The second approach accounts for the possibility of competing risks between the two events. In the data, when one failure event occurs, it precludes the other event from occurring, i.e. when a tariff is unilaterally removed, it cannot then experience a WTO dispute.[22] The first approach, using a Cox model that treats *Unilateral Removal* as an instance of right-censoring, is best when the risks of the two events are independent (Sueyoshi, 1992, pp. 30).[23] Theoretically, there are reasons to suspect that the two risks are not independent. For instance, if a country decided not to initiation a WTO dispute because it thought that the tariff was likely to be unilaterally removed, the independence assumption would be violated.

To account for this possibility, I also model the probability of the two events jointly, using a Bayesian multinomial probit (MNP) model from Imai and van Dyk (2005) which does not require any assumptions of independence among the risks.[24] The MNP also allows me to compare the effects of the explanatory variables on both risks, analyzing the direction and magnitude of certain variables on the risk of a *WTO Dispute* and *Unilateral Removal*.

Following Imai and van Dyk (2005), I let the observed multinomial variable, Y_{it}, take on a distinct value depending on the status of tariff i at time t. Let $j = 1, 2, 3$ index the 3 statuses, *WTO Dispute, Unilateral Removal, In Effect*. Call $j = 3$, *In Effect*, the base category. Let $W_{it} = (W_{it1}, W_{it2})$ be a vector of 2 latent variables, associated with *WTO Dispute* and *Unilateral Removal*, for tariff i at time t. The observed variable, Y_{it} is modeled in terms of W_{itj} via:

[22] It is technically possible that a country could initiate a WTO dispute over a unilaterally removed tariff, but this does not occur in reality. Similarly, the U.S. could unilaterally remove a tariff in response to a WTO dispute, but this would not occur via a negative final ruling or revocation.

[23] In the latent failure time approach to time-until-failure analysis, each observation, i, has a latent failure time, T_{ij}, for each of the j competing risks. We only observe the first failure, $min(T_1, T_2, ..., T_j)$, or failure due to the risk for which the latent failure time is the quickest. The independence assumption says that these latent failure times, the T_{ij}'s are conditionally independent of one another.

[24] The multinomial probit model is often associated with the concept of discrete choice, where an agent can choose from a menu of actions or options. Examples are voters choosing which candidate to vote for from a list or consumers choosing what brand of a good to purchase. The multinomial probit model is not limited to choice; it can also describe any situation where the model is of a dependent variable that can take on any of a number of distinct values. For analyzing categorical data, the MNP is often preferred to the multinomial logit (MNL) model because the MNP does not require an Independence of Irrelevant Alternatives (IIA) assumptions. The IIA assumption made in the MNL approach are very similar to the assumption of independence of competing risks in the time-until-failure approach.

$$Y_{it}(W_{itj}) = \begin{cases} 0 & \text{if } max(W_{it}) < 0 \\ j & \text{if } max(W_{it}) = W_{itj} > 0 \end{cases}$$

where $max(W_{it})$ represents the largest value in the vector W_{it}. The latent variables are modeled as a function of the k observed covariates.

$$W_{it} = X_{it}\beta + e_{it}, e_{it} \sim N(0, \Sigma)$$

X_{it} is a matrix of $2 \times k$ matrix of observed covariates and β is a $k \times 1$ vector of coefficients. $\Sigma = (\sigma_{lm})$ is a positive definite 2×2 matrix. For identification, the model assumes that $\sigma_{11} = 1$. The Bayesian approach implemented here uses the MCMC procedure developed by Imai and van Dyk (2005) to sample to sample from posterior distributions of β and Σ, based on particular prior distributions. I use very agnostic priors, where each element of β is distributed normally with mean 0 and variance 100.[25] For the main MNP model, I used a burn-in of 20,000 draws and kept every fourth draw from 70,000 subsequent draws.[26]

Results: Risk of a WTO Dispute

Table 3.1 shows the coefficients estimated from the above Cox model, using a series of model specifications.[27] The first model includes only the main explanatory variables and the retaliation variables: *U.S. Elec. Year*, *U.S. Unemployment*, their interaction, *U.S. Exports* and *U.S. Imports*. The second model adds variables describing Plaintiff-side dynamics: *Plaintiff PCGDP*, *Plaintiff Unemployment*, *Plaintiff Election* and the relevant interaction. The third and fourth models account for possible calendar year trends with a counter variable that begins at 1 for the first calendar month of the dataset. I also include the quadratic expansion of the counter.

[25] Setting the prior variance to 100 means that the prior distribution is very diffuse and unlikely to influence results.

[26] For the models with calendar month and age polynomials included as covariates (described below), I set the prior variance to 80, used a 15,000 draw burn-in, and kept every fourth draw from 60,000 subsequent draws.

[27] I used the *coxph* program in the Zelig package for R (Lam, 2007). The regressions use robust standard errors and the Breslow method for breaking ties.

The results support the theoretical predictions. During U.S. election years, increased unemployment lowers the risk of a WTO dispute.[28] Conversely, during non-election years, increased unemployment is weakly associated with a higher risk of a WTO dispute. This is consistent with the possibility that plaintiffs wait until non-election years to initiate WTO disputes. If the plaintiff knows that the U.S. is in an election year, and is more hostile to free trade, they are more willing to delay their WTO disputes for fear of not gaining concessions from the dispute, or worse, of provoking backlash.

Other theories receive mixed support. For retaliation, increased U.S. exports to the plaintiff are associated with a higher risk of a WTO dispute as predicted. But increased imports from the plaintiff, i.e. weakened plaintiff leverage, are also weakly associated with a higher risk of a WTO dispute. Tariffs against richer plaintiffs have a weakly higher risk of WTO disputes. Explanations based on plaintiff unemployment and electoral dynamics receive little support.

Results: Competing Risks

Table 3.2 reports summary statistics of the posterior densities for the coefficients in the multinomial probit specifications.[29] The top half of Table Table 3.2 reports the coefficients for the effect of the covariates on the probability of a *WTO Dispute* relative to the probability that a tariff remains *In Effect*. The bottom half reports the coefficients for the effect of the covariates on the probability of a *Unilateral Removal* relative to the probability that a tariff remains *In Effect*. A positive coefficient means that an increase in that covariate increases the probability of that event, relative to the base category (*In Effect*). I report the mean and 95 $\%$ confidence bands associated with each covariate's posterior density, for a *WTO Dispute* and for *Unilateral Removal*. To (greatly) ease interpretation,

[28] Recall that the "total" effect of unemployment accounts for the coefficient on the interaction term and the constituent terms. For example, the "total" coefficient for unemployment during an election year in Model 1 is $-5.44 + 0.088 = -0.456$.

[29] I use the same progression of models as in the Cox results, but in models with calendar month trends, I also add a quadratic age polynomial. The *Age* variable is a counter that begins at 1 for the first month that a tariff is *In Effect*. I also include *Age* squared. This is akin to the baseline hazard in the Cox approach.

I focus on the substantive effects of the variables of interest on the probability of a *WTO Dispute* and *Unilateral Removal*.

First, Figure 3.2 shows the effects of *U.S. Unemployment*, broken down by *U.S. Election Year*, on the probability of a *WTO Dispute*.[30] The pattern from the Cox regressions again is apparent in the multinomial approach. During election years, higher unemployment decreases the probability of a WTO dispute. Other countries are less likely to initiate WTO disputes against the U.S. during politically sensitive times when broader audiences are more supportive of protectionism. Conversely, higher unemployment increases the probability of a dispute during non-election years. During times of high unemployment, other countries delay their initiation of WTO disputes against the United States until less politically-sensitive time periods.

Second, Figure 3.3 shows the effects of *U.S. Unemployment*, broken down by *U.S. Election Year*, on the probability of *Unilateral Removal*. Importantly, the inter-electoral dynamics associated with the a WTO dispute are *not* present when considering *Unilateral Removal*. Regardless of election year, higher unemployment decreases the probability of *Unilateral Removal*. This is consistent with Hansen (1990) who finds that higher industry-level unemployment increases the probability of affirmative ITC rulings.[31] During times of higher unemployment, the U.S. is less likely to unilaterally remove its tariff barriers.

This finding is also an informal placebo test of theory proposed above. We would not expect the political economic effects of unemployment and electoral dynamics that affect the probability of a WTO dispute to also affect the decisions of bureaucrats who are making decisions over unilateral removal. Bureaucracies making decisions over affirmative or negative rulings are not elected officials making decisions in the shadow of a possible backlash from a broad constituency. While

[30] For the predicted probability figures, I use the *predict* command included in the *MNP* package. I drew $\sim$ 1,000 draws from the posteriors of each coefficient and calculated the probabilities based on a matrix of values for the covariates, generating predictions from each posterior draw. The Figures show the means of these predictions. I varied *U.S. Unemployment* from 4.5 to 5.7, which is are approximately the sample 25th and 75th percentiles. The continuous control variables were set to their sample means, with *Plaintiff Election Year* was set to 1. The vertical axes are the predicted probabilities *for a single month-long interval*, which is why the scale of these axes is small. For the predicted probability figures, I used the results from Model 7.

[31] Blonigen and Bown (2003) find a similar, though statistically insignificant, result.

bureaucratic agents are influenced political agents who control their purse strings, e.g. the chairs of the House Ways and Means Committee, those principals are beholden to more narrow constituent interests, rather than a broader reaction to a WTO dispute. It is encouraging for the results that inter-electoral dynamics are present only for *WTO Disputes*, and not for *Unilateral Removal*.

Further support comes from evidence regarding the relationship between the number of AD and CVD petitions filed and the overall U.S. unemployment rate. We would be worried if there was strong evidence that firms or the bureaucracies making decisions over AD and CVD petitions anticipated possible WTO disputes, potentially biasing the above findings. If firms filed fewer petitions in times of low unemployment and more petitions in times of higher unemployment, then that would be evidence that they possibly anticipated future WTO disputes, and resulting pro-free trade audience support. Figure 3.4 plots the number of new AD and CVD petitions against the U.S. unemployment rate. Fortunately, we do not find evidence of anticipatory behavior. There are not more petitions filed during times of higher unemployment.[32]

Third, Figure 3.5 and Figure 3.6 show the effects of U.S. exports and imports on the probability of a *WTO Dispute* and *Unilateral Removal*.[33] As the U.S. exports more to the country targeted by a tariff, the country is more likely to initiate a WTO dispute. Larger countries and countries to whom the U.S. exports more have greater leverage over the United States, and are therefore better able to compel the United States to comply with adverse WTO rulings, even when accounting for relevant political economic concerns. The U.S. is also more likely to unilaterally remove protectionist barriers against these countries. This is consistent with Blonigen and Bown (2003) who find that the potential for trade retaliation can deter U.S. antidumping activity.

[32] This result is the same if I break the Figure down by election year verses non-election years. This result should not be surprising. When making decisions over whether or not to file AD and CVD petitions, firms focus almost exclusively on their own situation, not overall economic conditions. Bown (2005*b*) models the decision over whether to file a petition and whether a WTO dispute results. He does not find substantively different results from models that do and do not account for the first stage, or selection decision- to file a petition.

[33] These predictions set the all other covariates to their sample means, with *U.S. Election Year* set to 1. The lines represent the mean of the predictions associated with 300 draws from the posterior coefficient densities.

The opposite is true of U.S. imports. As the U.S. imports more from a particular country, that country is less likely to initiate WTO disputes against the United States, because they have less leverage over the U.S. even if they were to win a WTO ruling against a protectionist barrier. The U.S. is also less likely to unilaterally remove protectionist barriers. This is consistent with existing work that finds that import surges and import penetration are an important impetus for antidumping and countervailing duty petitions (Irwin, 2004; Sabry, 2000; Busch, Reinhardt and Shaffer, 2009; Allee, N.d.).

Conclusion

If institutions, and dispute settlement in particular, are important alarms that can mobilize domestic audiences against defections from international agreements, then why do plaintiffs wait so long before sounding the alarm? Since countries often wait months or even years before initiating disputes over violations of international agreements, then explaining *when* they sound the alarm is as important as explaining *whether* they sound the alarm.

Accounting for features of the audience who hears the alarm can explain variation in the timing of disputes. If the goal of a dispute is to activate pro-compliance audiences to mobilize against offending defendant government policies, then the likely reaction of those audiences affects the expected value of a dispute for the plaintiff. Audiences who strongly support compliance make disputes more attractive, and politically powerful audiences are better able to influence their government's policies. On the other hand, audiences opposed to compliance or impotent audiences make disputes less attractive. Disputes should be most likely when domestic audiences are both willing and able to encourage their government to comply with its international obligations.

Data from WTO disputes against the United States is consistent with these predictions. Foreign countries are less likely to challenge U.S. tariffs when politicians face stronger audiences in favor of protectionist measures. These findings lend support to the argument that dispute settlement can be an important information transmission mechanism. But ultimately, the effects of this mechanism are circumscribed by the preferences and strength of the audience learning from this information.

The results also showed that informational/domestic audience explanations and power politics/retaliation explanations both affect member state behavior. Often, debates over whether or not “institutions matter” are cast in stark terms. On the one hand, institutions affect member state behavior because of a particular role for the institutions, e.g. information, credible commitments, audience costs, etc. On the other hand, evidence of institutional effects on member state behavior are sometimes thought to be artifacts of underlying power relations that govern member state interactions, regardless of institutional effects. Institutions appear to affect member state behavior, but only because of the power politics underlying international relations. These results show that the answer need not be one or the other. Rather, both dynamics can be at work. Just because “institutions matter” does not imply that “power politics” do not, and vice versa.

Figure 3.1: Months Between Tariff and WTO Dispute

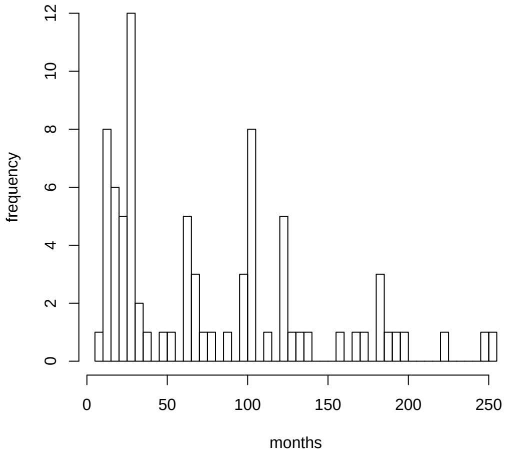


Figure 3.2: Effects of U.S. Unemployment on Pr(WTO Dispute)

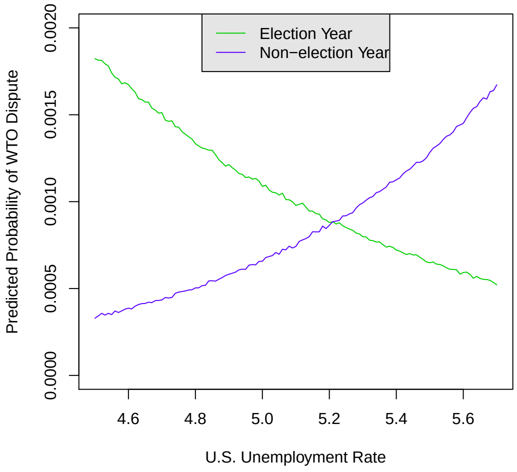


Figure 3.3: Effects of U.S. Unemployment on Pr(Unil. Removal)

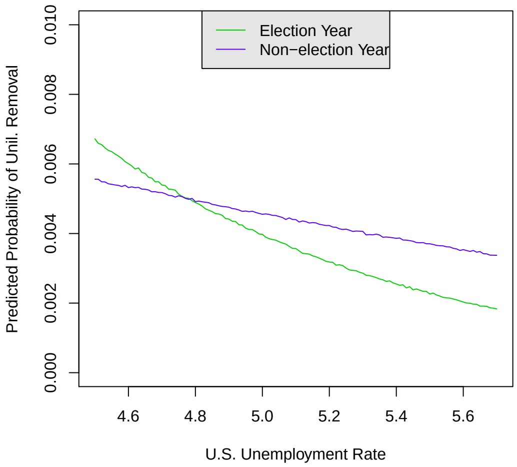


Figure 3.4: New Tariffs verses U.S. Unemployment

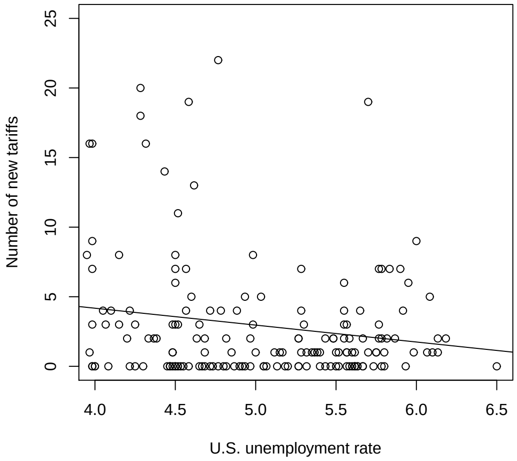


Figure 3.5: Effects of U.S. Exports on Probability of Exit

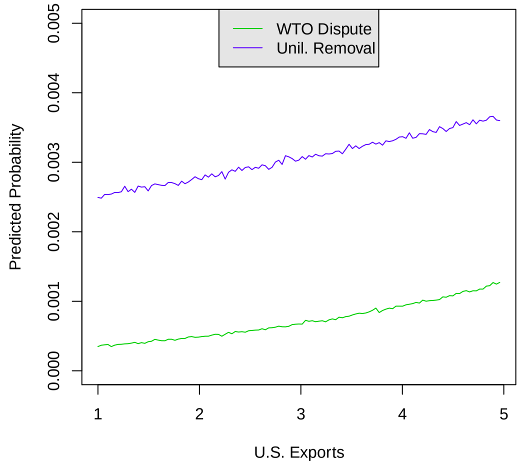


Figure 3.6: Effects of U.S. Imports on Probability of Exit

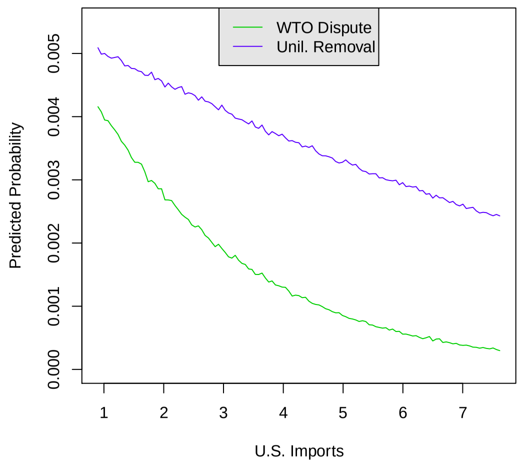


Table 3.1: Cox Models: Risk of WTO Dispute

	Model 1	Model 2	Model 3	Model 4
U.S. Elec. Yr. * U.E.	-0.544*	-1.978***	-0.521*	-1.394**
	(0.322)	(0.588)	(0.295)	(0.454)
U.S. Unemployment	0.088	1.025**	0.063	0.678*
	(0.127)	(0.367)	(0.162)	(0.319)
U.S. Elec. Yr.	3.237*	10.278***	3.251*	7.469***
	(1.587)	(2.960)	(1.502)	(2.364)
U.S. Exports	0.018	0.280***	0.025	0.267***
	(0.045)	(0.068)	(0.035)	(0.073)
U.S. Imports	0.009	-0.348***	0.011	-0.336***
	(0.038)	(0.075)	(0.031)	(0.087)
Pl. PCGDP		5.41x10^{-5}***		0.000***
		(1.35x10^{-5})		(0.000)
Pl. Elec. Yr. * U.E.		-0.006		-0.003
		(0.066)		(0.066)
Pl. Unemployment		-0.027		-0.021
		(0.033)		(0.031)
Pl. Elec. Yr.		0.341		0.297
		(0.518)		(0.544)
Month			0.071***	0.074***
			(0.016)	(0.021)
Month Sq.			-3.04x10^{-4}***	-3.56x10^{-4}***
			(7.84x10^{-5})	(1.15x10^{-4})
Log-likelihood	-404.609	-235.620	-386.981	-226.487
Num. Tariff	574	437	574	437
Num. Disputes	78	52	78	52

Coefficient estimates from Cox proportional hazards model with robust standard errors. *WTO Dispute* is the failure event, with *Unil. Remov.* treated as right-censoring.

Table 3.2: MNP Models: Risk of WTO Dispute

	Model 5	Model 6	Model 7	Model 8
		WTO Dispute		
U.S. Elec. Yr. * U.E.	-0.253	-0.973	-0.300	-0.785
	(-0.438, -0.067)	(-1.365, -0.606)	(-0.562, -0.043)	(-1.206, -0.392)
U.S. Unemployment	0.048	0.517	0.046	0.431
	(-0.048, 0.143)	(0.263, 0.813)	(-0.105, 0.176)	(0.135, 0.751)
U.S. Elec. Yr.	1.416	5.049	1.720	4.143
	(0.463, 2.332)	(3.186, 7.056)	(0.377, 3.026)	(2.163, 6.262)
U.S. Exports	-0.001	0.135	0.010	0.135
	(-0.027, 0.023)	(0.069, 0.212)	(-0.020, 0.043)	(0.073, 0.206)
U.S. Imports	-0.017	-0.167e	0.003	-0.169
	(-0.034, -0.002)	(-0.259, -0.086)	(-0.027, 0.031)	(-0.255, -0.093)
Pl. PCGDP		2.216 x 10^{-5}		2.277x10^{-5}
		(1.013 x 10^{-5}, 0.000)		(1.074x10^{-5}, 0.000)
Pl. Elec. Yr. * U.E.		-0.007		-0.005
		(-0.069, 0.053)		(-0.076, 0.059)
Pl. Unemployment		-0.014		-0.014
		(-0.054, 0.024)		(-0.062, 0.024)
Pl. Elec. Yr.		0.191		0.162
		(-0.303, 0.680)		(-0.365, 0.693)
Intercept	-3.216	-6.974	-5.009	-7.841
	(-3.958, -2.537)	(-8.689, -5.383)	(-6.120, -2.851)	(-9.867, -6.156)
		Unilateral Removal		
U.S. Elec. Yr. * U.E.	-0.240	-0.072	-0.142	-0.025
	(-0.368, -0.120)	(-0.224, -0.007)	(-0.423, -0.051)	(-0.121, -0.001)
U.S. Unemployment	-0.126	-0.033	-0.089	-0.016
	(-0.198, -0.058)	(-0.122, -0.003)	(-0.173, -0.036)	(-0.073, -0.001)
U.S. Elec. Yr.	1.168	0.3460	0.688	0.118
	(0.592, 1.782)	(0.032, 1.092)	(0.250, 2.032)	(0.007, 0.577)
U.S. Exports	0.021	0.010	0.015	0.003
	(0.005, 0.038)	(8.909 x 10^{-4}, 0.034)	(0.004, 0.039)	(6.982x10^{-5}, 0.017)
U.S. Imports	-0.017	-0.012	-0.011	-0.003
	(-0.034, -0.002)	(-0.039, -0.001)	(-0.031, -0.002)	(-0.019, 0.000)
Pl. PCGDP		3.202 x 10^{-06}		1.104x10^{-6}
		(3.955 x 10^{-07}, 0.000)		(8.808x10^{-8}, 0.000)
Pl. Elec. Yr. * U.E.		-0.007		-0.002
		(-0.025, 0.000)		(-0.001, 0.000)
Pl. Unemployment		0.002		8.103x10^{-4}
		(-8.424 x 10^{-4}, 0.011)		(-3.260x10^{-4}, 0.005)
Pl. Elec. Yr.		0.065		0.020
		(0.005, 0.230)		(-9.952x10^{-5}, 0.107)
Intercept	-1.694	-0.566	-0.955	-0.201
	(-2.110, -1.300)	(-1.547, -0.081)	(-2.214, -0.512)	(-0.896, -0.019)
Calendar Month Trends	N	N	Y	Y
Age Trends	N	N	Y	Y
Num. Tariff	574	437	574	437
Num. WTO Disputes	78	52	78	52
Num. Unil. Remov.	318	261	318	261

Mean of posterior density for each covariate, for *WTO Dispute* and *Unil. Removal*, with 95% confidence bands.

Base category is *In Effect*.

Chapter 4

Micro-level Evidence: Preferences Over Consistency and Policy

According to audience costs theory (ACT), audiences punish policymakers for committing to one policy and then reneging on that promise. In international relations research, this theory has been frequently applied to crisis bargaining and international cooperation. In the latter context, policymakers commit to certain policies when they negotiate, sign, and ratify international agreements or join an international institution. ACT predicts that audiences punish policymakers who choose noncompliant policies that contravene their international obligations. From the policymaker's perspective, these *ex post* audience costs facilitate cooperation by making compliance more attractive *ex ante*, and therefore make international agreements a more credible commitment (Leeds, 1999).

The key assumption of ACT is that audiences have preferences over *consistency*. Audiences care about whether a policymaker's actions are consistent with past promises. In his original conception of audience costs, Fearon (1994) argue that inconsistency creates the opportunity for domestic political opponents to criticize the incumbent for damaging the country's international "credibility, face or honor" (581). Smith (1998) argues that audiences punish inconsistency because breaking commitments signals a leader's incompetence. Ashworth and Ramsay (2009) derive conditions under which audiences impose costs for backing down on leaders as part of an optimal incentive scheme contracted between the leader/principal and the audience/agent. In the context of international law, legalized commitments are especially costly to break, because domestic audiences may “modify their plans and actions in reliance on such commitments” and because audiences often have a normative aversion to breaking the law Abbott and Snidal (1998, 428).

However, audiences also have preferences over *policy*. Audiences care about the actual policies that are implemented, regardless of their consistency with past statements. Consider the (stark) example of a worker who stands to lose her job if their elected representative lowers tariffs on certain imports. Even if those tariffs violate free trade agreements, the worker is unlikely to support a policy of lower tariffs. In other words, the worker’s preferences over policy (high tariffs preferred to low tariffs) trump her preferences over consistency (high tariffs are inconsistent with prior commitments, while low tariffs are consistent).

A similar divergence between preferences over consistency and preferences over policy arises in virtually every crisis bargaining and international cooperation context. A voter might have preferences over whether their leader follows through with deterrent threats, but the voter may also have strong preferences over whether her leader should pursue policies that entail threats or possible military action, irrespective of their consistency with past promises. International agreements often prescribe that member states make costly, though mutually beneficial policy adjustments. These adjustments tend to create winners and losers among voters. Whether a voter gains or loses from policy adjustments made in the name of international cooperation likely has a strong effect on her reaction to that policy, irrespective of whether those policies are consistent or inconsistent with her country’s international agreements.

This paper decomposes audience reactions to policymaker decisions over international cooperation into two components: a consistency effect and a policy effect. Decomposing consistency effects and policy effects is important for the theoretical and empirical evaluation of how international agreements and institutions affect member state behavior. If consistency effects are strong, as predicted by ACT, then this is a cause for optimism: audiences, because of their penchant for consistency, are powerful forces for compliance with agreements. However, if policy effects are important for audience reactions, then the effects of international institutions on member state policy are at least partially constrained by audience preferences over policy. Audiences may care about consistency, which creates a space for institutions and agreements to have an independent influence on member state behavior, but if policy preferences are too strong, then the effects of institutions and agreements are lessened.

To distinguish between consistency and policy effects, I embedded an experiment in a survey conducted in May of 2012. The survey consisted of two parts. The first part, the main experiment, presented respondents with a hypothetical situation regarding a policymaker's decision over whether to implement protectionist trade barriers. After respondents were given arguments in favor of (pro's) and opposed to (con's) the trade barriers and told about their policymaker's decision, they were asked whether they approved or disapproved of this decision. Treatment consisted of randomly assigning the con that respondents received, with one con pertaining to the consistency of trade barriers with previous international agreements. Similar to Tomz (2007, 2008) and Levendusky and Horowitz (2012), this part of the survey captures the effects of consistency on approval of policymaker decisions.

The second part of the survey asked respondents about their preferences over trade policy and also asks a set of questions shown to be predictors of voter preferences over trade policy. This allows me to examine whether, and to what degree, the respondent's preferences over trade policy moderate consistency effects. I can examine whether treatments based on consistency have a stronger or weaker effect depending on the respondent's predicted policy preferences.

As in previous studies, when looking at the entire sample of respondents, I find strong consistency effects. When respondents are told that their leader's policies were inconsistent with past promises, their approval of their leader's actions decreases significantly. However, unlike previous research, I show that this effect is only present for respondents who do not already hold strong policy preferences. For respondents with strong preferences over the policy in question, informing them of the inconsistency between their leader's policy and past agreements has a significantly smaller effect. Even for respondents without strong policy opinions, I show that placebo treatments are almost as strong as consistency treatments. Giving respondents a hollow, content-less reason to oppose a policy is almost as effective in triggering their disapproval as telling the respondent that the policy is inconsistent with past agreements.

These findings suggest that policy preferences are a stronger explanator of audience reactions to their leader's policies, while audience preferences over consistency are of secondary importance. As a result, leaders choosing policy are more constrained by the preferences of their audience than by their past commitments or international agreements. Institutions and agreements are likely to have weaker effects for countries with audiences who are hostile to the policies entailed in those commitments. They are also likely to have weaker effects over issue areas where audiences have the strongest preferences over policy. The implication of this is that a key challenge facing international institutions is to not simply provide information or awareness about leaders who violate their international obligations, but also to persuade stubborn audiences who do not necessarily support compliance with those obligations in the first place.

Consistency and Policy Preferences

Audience Costs Theory (ACT) argues that domestic populations punish leaders who make commitments to certain policies or courses of action and then choose policies that are inconsistent with those commitments (Fearon, 1994). Audience costs have alternatively been described as "the surge in disapproval that would occur if a leader made commitments and did not follow through," (Tomz, 2007, pp. 823) and "the punishments, in the former of lower support, meted out by domestic populations against leaders that make foreign threats but then ultimately back down" (Levendusky and Horowitz, 2012, pp. 324). The punishment is often thought of as electoral: voters are less likely to return promise-breaking leaders to office. Since policymakers make decisions in the shadow of this potential punishment, audience costs affect the credibility of a policymaker’s promises and commitments, and in turn, affect the calculus of other leaders interacting with that policymaker.

The implications of this theory have been applied to both the context of crisis bargaining and international cooperation. In crisis bargaining situations, country A makes a deterrent threat regarding country B, saying “If you (country B) do X, then we (country A) will do Y.” If country B does action X, and country A does not respond with action Y, then ACT hypothesizes that audiences in country A will punish their leaders for backing down. A deterrent threat made by a leader who is sensitive to these costs is thought to be more credible than a threat made by a leader who would not suffer audience costs.

Similar arguments abound in the context of international cooperation. Signing international agreements or joining international institutions helps leaders raise the *ex post* costs of defecting from an agreement.[1] ACT hypothesizes that leaders who break their international agreements will suffer audience costs, which can make compliance with an agreement more attractive than defection. The prospect of this audience punishment creates a strong disincentive for a leader contemplating policies that do not comply with international obligations.

At its core, ACT is thus a theory of audience preferences over consistency between words and deeds. But audiences also undoubtedly have preferences over the deeds or actions themselves, irrespective of their consistency with past actions. An audience member assessing their leader’s performance in the context of international cooperation might care about the consistency of their leader’s promises and policies, but they also have preferences over the actual actions of their leader. Cooperation occurs when states agree on mutually beneficial policy adjustments that they would not have otherwise implemented unilaterally (Keohane, 1984). These policy adjustments impact audience members differently, creating winners who benefit from the policy adjustments and losers who do not. Trade policy adjustments have distributional impacts- raising and lowering tariffs, increasing or decreasing subsidies, or changing monetary policy benefits some

[1] For a more extensive review of this argument, see Simmons (2010).

audience members at the expense of others. In factor endowments theories of trade, tariffs are thought to harm owners of abundant factors and benefit owners of scarce factors, as hypothesized by the Stolper-Samuelson Theorem. In specific factors theories of trade, tariffs benefits and harm workers in different sectors or industries. Exactly *who* wins and loses depends on the particular economic model, but the presence of winners and losers is a common feature. The perceived or actual effects of trade policy adjustments have been linked to support or opposition to policies and candidates as well as the political cleavages that arise regarding trade policy (Rogowski, 1987; Hiscox, 2002). Milner and Tingley (2011) find that legislative voting patterns on trade policy bills are consistent with political economic predictions derived from the Stolper-Samuelson Theorem. Margalit (2011) finds that job-losses from offshoring had a significant effect on voter support for incumbents between 2000 and 2004. While these studies were not designed to test preferences over consistency, their findings regarding the behavior of legislators and voters are supportive of theories about preferences over policy.

In virtually every issue concerning international cooperation, there are groups within countries who support the policies proscribed by agreements and institutions and groups that oppose them. For instance, a rich body of literature examines variation in support for European integration both across and within countries.[2] A similar body of literature examines variation in domestic political support for international cooperation on climate change and the environment.[3]

In the highly-charged context of human rights and war crimes, there is significant variation within countries over whether to support compliance with international agreements. Compliance with these agreements often involves condemnation and punishment of recently removed leaders or even of current elected officials. An audience member's support for the politician or governing group being accused of human rights violations strongly tempers her preferences over whether to that politician or group should be punished. In 2005, the government of Kenya ratified the Rome Statute, which exposed Kenyan nationals to prosecution by the International Criminal Court

[2] For a survey of these theories, see: Gabel (1998).

[3] For a recent example, see: Kelemen and Vogel (2010).

(ICC) should they commit war crimes, crimes against humanity, or genocide. During and after the 2007 presidential elections, violence broke out between supporters of the incumbent, whose strongest support came from the country’s central and eastern regions, and the opposition, whose support came primarily from the western regions. In March of 2011, the ICC began the indictment process against six politicians, from both of the incumbent and opposition’s political parties, for their alleged roles in the post-election violence.

In January of 2012 a nationally administered poll asked “Are you happy or unhappy that the Hague/The ICC is pursuing the six suspects of post-election-violence?” Support for the ICC varied significantly across regions. 82% of respondents in the western region of Nyanza answered that they were happy with the ICC. In the Central region, only 44% of respondents answered that they were happy with the ICC.[4] It is highly likely that this variation is driven by preferences over policy, not preferences over consistency. Unsurprisingly, support for the ICC was largely driven by the region’s underlying support for particular politicians who had been indicted. In regions where indicted politicians enjoy significant public support, the public is much less supportive of the ICC process. In regions that perceive the ICC as a way to punish unpopular out-group politicians, the ICC process receives stronger support. Far from uniting the country under the ICC’s goal of ending impunity for crimes against humanity, the ICC’s actions have polarized the country, according some analysts, increasing divisions between communities supporting or opposing indicted politicians.[5]

Even in the canonical ACT context, crisis bargaining, audience members have strong policy preferences. Audiences care about the decision to issue compellent threats in the first place and about whether to use military force when the foreign country defies those threats. The act of unilaterally issuing a compellent threat in the first place is more than mere words. It signals the possibility of military action, however remote, and is an inherently coercive approach to foreign policy. A

[4] Survey conducted by South Consulting in January of 2012. See $http : //www.dialoguekenya.org/docs/KNDRFinalReportJanuary2012.pdf$ for the Draft Report.

[5] See: Rothmyer, Karen. “The International Criminal Court on Trial in Kenya.” The Nation. May 28, 2012. $http : //www.thenation.com/article/167810/international - criminal - court - trial - kenya.$

fundamental disagreement between so-called "hawks" and "doves" is over the best way to achieve foreign policy objectives: coercion verses persuasion, unilateral verses multilateral. Audience members also undoubtedly have preferences over whether to follow through with threats militarily. After all, the costs of military action may be large enough to persuade an audience member that backing down is the correct course of action. In their critique of ACT, Snyder and Borghard (2011) are skeptical of audiences who care more about consistency than policy substance, arguing in favor of a characterization of ACT that they attribute to Kenneth Schultz (2001): "publics are expected to punish leaders who back away from threats only if they agree with the threats on substantive grounds" (pp. 440).

Micro-level Evidence of Audience Costs

The two most well-known empirical studies of the micro-foundations of audience costs (Tomz, 2007; Levendusky and Horowitz, 2012) were in the context of crisis bargaining. In both studies, survey participants are told about an international crisis where one foreign country, the aggressor, is thinking about invading its neighbor country. In the treatment group, participants are told that the United States' leader threatened military action against the aggressor if it invaded; the aggressor invaded; and the United States did not follow through with its threat, refraining from military action while the aggressor invaded its neighbor. In other words, the treatment group is told that their leader's words and deeds were inconsistent. Participants assigned to the control group are told that the aggressor is thinking about invading its neighbor, but the United States' leader elects to stay out of the crisis- implicitly neither threatening nor using military action- and the aggressor proceeded with the invasion. All participants are then asked whether they approve or disapprove of the president's actions. As predicted by audience cost theory, approval is lower in the treatment group.[6] The treatment effects in both studies are large and significant. In Tomz (2007), respondents

[6] The two studies also embed other treatments to examine what factors moderate the degree to which audiences punish leaders for inconsistency. Tomz (2007) analyzes whether international factors, like the level of escalation or the predicted amount of U.S. casualties involved with following through on the threat, affect the magnitude of audience costs. Levendusky and Horowitz (2012) analyze whether domestic factors, like the party of the president and

who are told that their president's commitments and actions were inconsistent were approximately 16% more likely to disapprove of their president. In Levendusky and Horowitz (2012) respondents were approximately 22% more likely to disapprove of presidents who broke their commitments.

With this approach, there are two differences between the treatment and control groups- one pertaining to consistency and one pertaining to a potentially important policy decision. The first difference is the one desired by the investigators designing the survey. Survey respondents learn that the president is guilty of commitment-policy inconsistency in the treatment scenario, but not in the control scenario, which can affect their approval of the president. But the treatment also consists of a second difference- learning that the president *threatened* the aggressor country in the first place and then chose not to use military action, both of which are nontrivial policy decisions that could affect respondents' approval levels.

To see why preferences over policy could affect approval apart from consistency effects, consider two archetypal audience members: a "hawk" and a "dove." A "hawk" respondent is not averse to their president making threats and may also have a penchant for subsequent military action. If told that the president threatened but took no action, the "hawk" may disapprove because they preferred military action, irrespective of their preferences over commitment-policy consistency. A "dove" respondent may strongly dislike both threats to use force and military action. If told that the president threatened and backed down, they may disapprove because of their dislike of threats. This difference between treatment and control groups creates the possibility that disapproval stems from the respondent's dislike of inconsistency, dislike of policy, or both.

A third study, from Tomz (2008) uses an approach more closely resembling the one used here. Tomz (2008) analyzes results from a survey where respondents are first told about a situation involving whether to impose an embargo on goods imported from Burma into the United States. Respondents were randomly assigned different combinations of arguments for or against the embargo. The arguments in favor of the embargo (pro's) were that it would help human rights, or that

Congressional majorities match or the justification given by the president for backing down, affect the magnitude of audience costs.

it would help the U.S. economy. The arguments against the embargo (con's) were that it would hurt the Burmese economy, or that it would violate international law. The advantage of the approach used in Tomz (2008) is that treatment and control groups are not given different treatments with regards to policy choices- they aren't told of any policy choices at all. Rather, they are asked "How good of an idea is it for the United States to prohibit trade with Burma?" Respondents who were told that the embargo violated international law were 17% more likely to oppose the embargo than respondents who did not receive this argument. When told that a policy was inconsistent with prior obligations, respondents were significantly less likely to approve of that policy, and by a (relatively short) logical leap, would also be less likely to approve of a politician who chose that policy.

The approach used in this study also resembles Americanist work examining how voters respond to candidates who reposition, i.e. change their stance on an issue. Tomz and Van Houweling (2012) conduct survey research to analyze *valence* and *proximity* effects of candidate repositioning on voter opinions. Valence refers to characteristics that voters might find favorable in a candidate, like honesty, loyalty, and a commitment to keeping one's word. Similarly to the effect posited by audience costs theories, candidate repositioning negatively affects voters perceptions of the candidates along a valence dimension. But repositioning also has a proximity effect. Repositioning might bring the candidate closer to or further away from the voter's most preferred policy on a certain issue. A candidate who moves closer to the voter's ideal policy might suffer negative valence effects, but benefit from positive proximity effects. Tomz and Van Houweling (2012) use survey experiments where voters read about candidates' positions on taxes and abortion over time to determine the relative magnitudes of valence and proximity effects. They find strong evidence of valence effects, but these effects are more moderate for voters that care a lot about the issue at hand. Voters for whom the policy issue is more important care less about valence effects than voters who do not feel as strongly on the issue.

Experimental Design and Hypotheses

When audience members learn that their leader has chosen a policy that is inconsistent with previous international commitments, how much of their disapproval stems from their dislike of inconsistency and how much stems from their preferences over the particular policy chosen?

I embedded a randomized experiment in an online survey conducted in May of 2012. Survey respondents were recruited using Amazon's Mechanical Turk (mTurk) service and were directed to an external survey site to answer questions programmed with Qualtrics. mTurk provides access to a recruitment pool for survey respondents that is low-cost and comparable to nationally representative surveys. Berinsky et al. (2012) show that subjects recruited on mTurk are more representative of the U.S. population than convenience samples, though marginally less representative than subjects recruited via nationally representative internet-based samples or national probability samples. They replicate existing studies using subject pools recruited from mTurk and find results that are comparable to results produced with other subject pools.[7]

For the main experiment, respondents were presented with a hypothetical situation involving a fictional U.S. company, called *Arena Inc*. This company manufactured metal brackets, which, as respondents were told, U.S. construction companies used in building construction. Respondents were then told that a European company had recently begun producing similar brackets at a lower price, and that U.S. construction companies had begun buying the foreign brackets instead of the United States-produced brackets. I left the country unspecified to avoid tainting responses with the respondent's opinion of a particular country, and used a non-descript name, *Chapman Inc.*, for the foreign company. I chose to specify the European continent to avoid the risk that responses were influenced by the respondent's perceptions of the United States' most politically charged import partner- China. Respondents were then told that the president had to decide whether to impose a

[7] I owe appreciation to Peer et al. (2012), who provide a useful script for ensuring that mTurk workers do not take the survey more than once.

policy restricting imports of foreign-made brackets, and that "analysts" had lobbied the president in favor of and opposed to import restrictions.

Each respondent then received a standard pro-import restriction argument: "Some analysts have lobbied the president *in favor* of restricting imports of metal brackets from Europe. They argue that when U.S. construction companies buy foreign-produced brackets, Arena Inc. will be forced to lay off some of its employees." The treatment consisted of random assignment of one of three con's, i.e. arguments opposing import restrictions, or a null treatment, i.e. the respondent was not given a con. The text of the three cons is given below:

- International Law Treatment: Some analysts have lobbied the president *against* restricting imports of metal brackets from Europe. They argue that import restrictions violate free trade agreements between the U.S. and Europe, and Europe would sue the U.S. at the World Trade Organization.
- Economic Treatment: Some analysts have lobbied the president *against* restricting imports of metal brackets from Europe. They argue that when U.S. construction companies have to buy more expensive U.S. brackets, construction companies are forced to lay off some of their employees.
- Placebo Treatment: Some analysts have lobbied the president *against* restricting imports of metal brackets from Europe. They argue that such restrictions would have adverse consequences and that the benefits of the restrictions do not outweigh the costs involved in the measures.

The international law treatment captures the concept of consistency. The key content in the treatment is that import restrictions are contrary to a previous commitment, namely a free trade agreement. And this inconsistency would likely result in legal action against the United States. I incorporated the likelihood of legal action at the WTO to emphasize the rule of law and adjudication component of international agreements- namely that, when a country violates its agreement, a supra-national judicial body can be called upon to condemn those defections.[8]

The argument in favor of import restrictions most commonly invoked by politicians is the restrictions will help save jobs, as contained in the pro-import restriction argument that each respondent received. The economic treatment captures the notion that a policy of import restrictions might help save some jobs, but would also likely cost other jobs. I chose an argument pertaining to "downstream" jobs to match the pro-import restriction argument that every respondent received, which pertained to "upstream" jobs.

The placebo treatment matches the other two treatments in word count and structure, but does contain any specific content. Rather, it alerts respondents to some unspecified reason to oppose import restrictions. It is possible that respondents simply count the number of pros and cons when assessing a particular policy, so having any arguments listed as a con increases disapproval, regardless of the content of the treatment. Comparing the effects of the placebo treatment with the international law and economic treatments effects helps isolate the *additional* effect on approval that occurs because of the specific content of those treatments. As mentioned above, the null treatment consisted of not giving the respondent any of these three con arguments. To avoid stacking the deck in favor of finding effects for any one of the treatments, they each have identical sentence structures as well as very similar word counts and word tones.

After receiving the standard pro-import restriction argument and one of the four treatments (the three listed above or the null treatment), respondents were told that the president decided *in favor* of imposing import restrictions. Respondents were then asked if they approved or disapproved of the way the U.S. president handled the situation, and could answer: "Strongly Approve," "Somewhat Approve," "Neither Approve nor Disapprove," "Somewhat Disapprove," or "Strongly Disapprove." Respondents who answered "Neither Approve nor Disapprove," were then asked if

[8] The international law treatment is not meant to capture *why* the respondent might disapprove of violating an international agreement- reputation, respect for law, updating about leader quality, etc. Experimental tests of why audiences disapprove of leaders who break international agreements would be a fruitful area for future research.

they “leaned towards” approving or disapproving. This creates a six-point scale for approval of the president’s actions. This approval scale and wording closely resemble that of Tomz (2007) and Levendusky and Horowitz (2012).[9] I constructed a binary variable measuring approval verses disapproval which is coded 1 for respondents who answered “strongly/somewhat approve” or “lean towards approving,” and 0 otherwise. This variable measures approval rates, or the proportion of respondents who indicated that they approved of the president’s actions.

The key question of ACT is whether learning that a leader’s chosen policy is inconsistent with prior obligations decreases respondents’ approval of leaders who enact that policy? ACT predicts a negative treatment effect for the international law treatment. When respondents are told that their leader has chosen a policy inconsistent with a prior treaty, they should be more likely to disapprove of that leader’s policy choice, compared to other treatments. The null treatment provides a useful baseline, because I can compare approval levels for the three non-null treatments against approval levels for the group that received no “actual” treatment. I can also compare the relative magnitudes of the three positive treatments. Does learning that a policy was inconsistent with prior obligations decrease approval more than learning that a policy might harm certain domestic jobs? How much of this effect comes from the specific content of the treatment (international law verses economic), and how much comes from the fact that there respondent was given simple words on the page that were opposed to the policy (placebo treatment)?

The overall structure of the survey was as follows. Before the main experiment, I asked a series of questions about the respondent such as their age, sex, marital status, and state of residence. Respondents then read the hypothetical story described in the main experiment, the pro’s and con’s entailed in their randomly assigned treatment group, and answered the approve/disapprove questions. Respondents then answered a series of opinion questions and demographic questions. They first answered a series of five political knowledge questions that measured their familiarity with

[9] The only difference is that, unlike Tomz (2007), I did not allow respondents to indicate that they did not “lean towards approving or disapproving.” Levendusky and Horowitz (2012) did not ask the “lean towards” follow-up question.

certain world events. These were factual, multiple-choice questions with one correct answer.[10] I then asked a series of questions designed to measure the respondent's preferences over isolationism. Specifically, I asked "Agree or Disagree" questions pertaining to the U.S. role in the world, such as "The US government should just try to take care of the well-being of Americans and not get involved with other nations; Agree or Disagree." I then asked a series of questions measuring the respondents' levels of ethnocentrism, or the degree to which respondents perceive members of their own racial or ethnic in-group more favorably than out-group members.[11] I also asked a series of standard demographic questions, such as the respondent's party, ideology, income, education, etc. Apart from standard demographic questions, I asked questions related to empirical work on preferences over trade policy. I asked the respondents to estimate the current U.S. unemployment rate, as per sociotropic explanations (Mansfield and Mutz, 2009). I also asked whether the respondents were currently employed and whether they or a family member had ever been a member of a trade union.

The goals of the post-experiment questions were two-fold. First, asking these questions allows me to check that treatment assignment was not significantly correlated with any particular feature of the respondent, which might have affected the effect of the treatment on the respondent. I used logit regressions to ensure that observable respondent characteristic and responses were not significantly correlated with the probability of being assigned to a particular treatment group. For each of the four treatment groups, I regressed a dummy variable indicating that the respondent received that treatment on the respondent's age, gender, race, marital status, education level, political knowledge level, isolationism score, ethnocentrism score, employment status, income level, party, ideology, and union membership. The results are displayed in Table 4.1.

[10] Respondents were asked which party currently controlled the U.S. House of Representatives (Republicans), which country recently ousted Muammar Gaddafhi from power (Libya), who was the current Supreme Court Chief Justice (Roberts), which country was not a permanent member of the United Nations Security Council (India), and which country was not a member of the Allies during World War II (Switzerland)?

[11] The isolationism and ethnocentrism questions are identical to those used in Mansfield and Mutz (2009). I also standardized these responses in the same way as Mansfield and Mutz. The ethnocentrism and isolationism questions are standardized to have a mean of zero, with higher numbers indicating increased isolationism and ethnocentrism.

For each treatment group, I cannot reject the null hypothesis that the coefficients on these variables are jointly 0. The χ^2 statistics and associated p-values for each treatment group are: international law- 21.53 (p=0.253), economic- 19.55 (p=0.359), placebo- 16.11 (p=0.585), and null- 14.52 (0.695). Only a few respondent characteristics were singularly significant for particular treatment groups and none had strong substantive effects on the probability of particular treatments. These null results also obtain when I regress treatment only on characteristics that were elicited pre-treatment. The results are displayed in Table 4.2. The χ^2 statistics and p values are even lower in these regressions: international law- 6.82 (p=0.557), economic- 2.46 (p=0.963), placebo- 3.89 (p=0.867), and null- 7.83 (p=0.450). The only pre-treatment covariates that were significant in any regressions were that slightly more males received the international law treatment, and Asian respondents were slightly over-represented in the null treatment group relative to respondents who selected "Other" for their race.

The second goal of asking these post-experiment questions is to compare the relative magnitudes of consistency effects and policy preference effects. If ACT is correct, then the respondents' preferences over trade policies like import restrictions should not moderate consistency effects. In other words, learning that a policy is inconsistent with previous commitments, as in the international law treatment, should have the same effect for respondents who support import restrictions, oppose import restrictions, or do not feel strongly either way.

If, on the other hand, policy preferences are important, then we should see different international law treatment effects depending on whether the respondent supports or opposes restrictions on free trade. Respondents who strongly support import restrictions should care less that import restrictions are inconsistent with previous commitments. For these respondents, the international law treatment "pulls against" their preferences. In the absence of the international law treatment, some unknown factors underlie the respondent's support for import restrictions. The international law treatment must overcome these factors to move the respondent to disapprove of import restrictions.

Respondents who strongly oppose import restrictions should also show weakened international law treatment effects. For various reasons, these respondents already have a low approval level of import restrictions, so the international law treatment is just "yet another" reason to oppose a policy that they already oppose. Learning that import restrictions are inconsistent with previous commitments just moves anti-import restriction respondents closer to their "floor" level of approval. Respondents with strong preferences over trade policy should also be less susceptible to the placebo treatment. These respondents' preferences over trade policy are likely to be founded upon something stronger than hollow words. Giving these respondents a treatment with no content or new arguments should not have any significant effect on their level of approval or disapproval.

To measure policy preferences, I also asked a standard free-trade question in the middle of the lengthy set of post-experiment questions.[12] Specifically, respondents were asked: "As you may know, international trade has increased substantially in recent years. This increase is due to the lowering of trade barriers between countries, that is, tariffs or taxes that make it more difficult or more expensive to buy and sell things across international borders. Do you think government should try to encourage international trade or to discourage international trade?" Respondents could answer that government should try to "Encourage [free trade] a lot," "Encourage a little," "Neither encourage nor discourage," "Discourage a little," or "Discourage a lot."[13] I call respondents who answered that the government should encourage free trade either a little or a lot as pro-free trade respondents. Respondents who answered that the government should discourage free trade either a little or a lot are called protectionist respondents. Respondents who answered neither encourage nor discourage are called no opinion respondents.

Since this question was asked after the main experiment, I checked for evidence that treatments from the main experiment "contaminated" respondents' answers to the free trade question. The

[12] I asked the free-trade question after the political knowledge and isolationism questions in order to distance this question from the main experiment, but before the ethnocentrism and demographic questions to avoid priming their responses with concepts contained in the demographic questions. The half of the respondents who were not asked about free-trade were asked a benign question: "How often do you read the newspaper each week?"

[13] The framing and response set for this question are identical to that used by Mansfield and Mutz (2009).

survey was designed to dampen such effects by placing all of the political knowledge questions and isolationism questions between the main experiment and trade policy question. There is not strong evidence that the treatment received by each respondent affected their response to the free trade question. I used an ordered logit regression to estimate the effects of treatment assignment on free trade responses. I coded pro-free trade respondents as 1, no opinion respondents as 2, and protectionist respondents as 3, and regressed this variable on dummy variables indicating treatment assignment. The results are presented in Table 4.3. None of the treatment assignments had a significant effect on the probability of a respondent being pro-free trade, protectionist, or having no opinion. Being assigned to the international law treatment group did increase the probability of a respondent giving a more pr-free trade answer, relative to the null treatment group, but this effect was small and statistically insignificant. A χ^2 test also fails to reject the null hypothesis that the effect of treatment assignment on trade preferences is collectively zero. The likelihood ratio χ^2 statistic is 2.09 with an associated p value of 0.55.[14] The results are robust to multinomial regressions or difference in means tests for trade policy responses across treatment groups.

To check that respondents actually received the desired treatment, I asked them to recall the pro- and con- arguments that they had received in the main experiment from a list of four possible arguments. 86.4% were able to correctly recall that they had been given a pro-import restriction argument pertaining to layoffs by the U.S. metal bracket firm, among a list containing the correct answer and four fabricated arguments in favor of import restrictions. 62.2% were able to correctly recall the anti-import restriction that they had been given (if any) from a list containing each of the four possible treatments. The placebo treatment, unsuprisingly, was the weakest, with only 47.8% of respondents correctly recalling it. The international law and economic treatments were stronger, with 68.1% and 69.3% correctly recalling the con arguments that they'd been given.

[14] Originally, I randomly selected half of the respondents to receive this question. In the case that treatment assignment *was* affecting respondents answers to the free trade question, I wanted to use the half of the respondents who did answer that question as a "training dataset" to generate a model that estimated respondents' free trade preferences as a function of other covariates, with the goal of conditioning treatment effects on respondents' predicted free trade preferences. When the initial set of surveys displayed very little evidence that treatment assignment affected responses to the free trade question, I began asking all respondents the free trade question.

63.7% of respondents who received the null treatment correctly recalled that they had not been given an anti-import restriction argument. Both the pro- and con- manipulation check results were easily able to reject the null hypotheses that respondents guessed at random, i.e. that the proportion of correct responses was 0.25, at the 0.01 level.[15]

Experimental Results

Before examining comparing preferences over consistency and policy, I first present evidence of consistency effects that are analogous to existing studies (Tomz, 2007; Levendusky and Horowitz, 2012). Figure 4.1 and Table 4.4 show the percentage of respondents who approved of presidents who implemented import restrictions across each of the treatment groups. Among those who received the null treatment, 68.7% approved of the president's actions. Among those receiving the international law treatment, only 56.0% approved of the president's actions. The difference between the null approval rates and the international law treatment approval rates is an initial approximation of consistency effects. This difference measures the drop in approval that occurs when respondents learn that their president's actions violated prior agreements. Approval rates are 12.7% lower in the international law treatment group than in the null group. This difference is highly statistically significant (p value for the difference in means is < 0.01).[16][17]

The other two treatments do not seem to have any significant effects on approval rates. Among those who received the economic treatment, approval decreased slightly, relative to the null group, to 67.1%. Even direct economic concerns, like the possibility of job loss in other industries, does

[15] The null hypothesis is rejected in binomial tests that the proportion of correct answers is greater than 0.25 as well as simple difference in means tests.

[16] p values use the normal approximation of the Bernoulli data. The number of respondents in each group is much larger than traditional minimum values necessary to use the normal approximation. In the future, I will use the Bayesian Jeffrey's prior approach.

[17] The survey software allows researchers to record the amount of time the respondent spent on each page of the survey. I discarded results from respondents who spent less than 5 seconds reading the hypothetical scenario described in the main experiment or who spent less than 3 minutes on the entire survey. The average survey time, excluding some outliers who restarted the survey after initially stopping, was approximately 9.5 minutes. Similarly, respondents spent a little over 1 minute reading the text of the main experiment.

not appear to influence approval rates. Among those who received the placebo treatment, 64.4% approved, a 4.2% drop compared to the null group. Neither of these differences is significant at conventional levels.

These initial results appear to be a strong confirmation of ACT. The consistency between words and deeds appears to be the only factor with a significant effect on approval rates. However, the effect of consistency on approval is significantly moderated when broken down by respondent preferences over free trade. Figure 4.2 shows the approval rates for the international law treatment compared to the null treatment, broken down by whether respondents said that government should encourage, discourage, or neither encourage nor discourage free trade. These results, as well as the difference in approval rates with the null treatment and approval rates with the international law, economic and placebo treatments are shown numerically in Table 4.5.[18]

For pro-free trade respondents and protectionist (anti-free trade) respondents, the difference between approval rates for the null group and the international law group are small and insignificant. Among pro-free trade respondents, the approval rates in the international law treatment group were 49.6% compared to 55.0% for the null treatment group. The difference, -5.4% is less than half as large as the difference found for the entire sample (-12.6%), and is statistically insignificant (p value = 0.401). Among protectionist respondents, the approval rates for the international law treatment group were 89.7% compared to 95.8 for the null treatment group. Substantively, this difference of -6.1% is comparable to the effect for the pro-free trade group and is also insignificant (p value = 0.270).

The treatment effect found in the full sample is very strongly driven by respondents with no preferences over free trade. Among respondents who neither supported nor opposed free trade, the approval rates in the international law group were 52.2%, compared to 72.7% for the null group. The difference of -20.6% is substantively large and statistically significant (p value = 0.033).

[18] p values in this table also use the normal approximation of the Bernoulli data, but some of the cells are close to minimum values for this approximation to be appropriate. For now, I am retaining the normal approximation for convenience but the sample sizes are growing as more respondents take the survey.

Consistency effects are most strongly displayed for respondents without strong policy preferences, and consistency effects are much weaker for respondents who have an expressed opinion over the policy at hand. Learning that import restrictions were inconsistent with past obligations was unpersuasive for both free-trade and protectionist respondents. Neither group significantly decreased their approval rates when they learned that import restrictions violated free trade agreements. Learning that import restrictions violated free trade agreements only had a significant effect on respondents who did not hold strong opinions over free trade in general. Put simply, if the respondent felt that free trade was good, then learning that import restrictions were illegal had little effect, since it only reinforced this opinion. If the respondent felt that free trade was bad, then learning that import restrictions were illegal was insufficient to overcome the factors that drove their underlying aversion to free trade. Respondents without strong opinions on free trade were the most malleable, and most influenced by inconsistency between words and deeds.

These results are consistent with Tomz and Van Houweling (2012) analysis of domestic tax and abortion policy. They find that valence (consistency) effects are strongest among respondents who do not consider the issue to be very important. Among respondents who considered tax or abortion policy to be particularly important, proximity (policy) effects were most important. If the respondent cared strongly about the issue, then their support for a political candidate was driven less by the candidate's consistency on the issue and more by the respondent's expectations about the policy that candidate would choose.

This pattern is also displayed when considering the economic and placebo treatments. Respondents with established opinions on free trade were less moved by either treatment. Respondents without strong opinions on free trade were more influenced by both treatments. Figure 4.3 shows the approval rates for the placebo group compared to the null group, broken down by the respondent's free trade preferences. Figure 4.4 does the same for the economic treatment group compared to the null group. The economic treatment actually has a positive (though small and insignificant) effect on approval rates among pro-free trade respondents, 0.8%. It has a larger and negative effect among no-opinion and protectionist respondents, -8.2% and -5.8% respectively, though both are insignificant.

Among pro-free trade and protectionist respondents, the difference between approval rates in the placebo and null groups were very small and insignificant. Respondents with stronger views on free trade were very weakly affected by receiving the placebo treatment. For pro-free trade respondents, the placebo treatment decreased approval relative to the null group by only -2.1%. For protectionist respondents, the placebo treatment effect was only -4.7%. Yet for respondents expressing no opinion, the placebo treatment managed to decrease approval by -8.8%, though this difference falls short of conventional significance levels (p value = 0.314).

The strength of the placebo treatment for respondents without strong policy opinions suggests that the effect of the international law treatment may have as much to do with simply treating respondents with *any* con- argument as it does with the specific content contained in the international law treatment. In other words, limiting our analysis only to the respondents where we found a significant international law treatment effect, the international law treatment effect was statistically indistinguishable from the placebo treatment effect. In his 2008 study, Mike Tomz distinguishes these two effects as “addition” and “substitution” effects. Addition effects arise when the respondent is given an additional reason to approve or disapprove of a leader’s actions. Both the international law and placebo treatments have an addition effect relative to the null treatment, since the respondent receives no con arguments with the null treatment. Substitution effects arise when comparing approval rates, substituting one argument for another. In this case, substituting the international law treatment for the placebo treatment decreases approval an additional -11.7% relative to the null treatment. But this difference is statistically insignificant (p value = 0.225). While we can confidently say that both the international law and placebo treatments have additive effects, we cannot confidently say that the international law treatment has substitutive effects relative to the placebo treatment.

The results overall suggest that preferences over policy are a stronger driver of leadership approval than preferences over consistency. To predict a respondent's approval of a leader who implemented import restrictions, the respondent's preferences over the policy of import restrictions is a better predictor than whether or not the respondent knows that the policy is inconsistent with the leader's previous commitments. Using simple OLS, regressing the respondent's approval on dummies indicating which treatment the respondent received yields a small R^2 value of 0.0101. Regressing approval on the respondent's expressed preferences over free trade, however, yields an R^2 value 0.0924, increasing the explained variation in approval by a factor of approximately 9. Logit regressions yield similar pseudo-R^2 values of 0.0845 and 0.0076 for policy effects and consistency effects respectively. The AIC and BIC are much lower for the logit policy effects regression, 1083.983 and 1098.386, than for the consistency regression, 2175.005 and 2196.694.[19]

Conclusions and Broader Implications

Audience costs theories predict that voters impose substantial punishment on leaders whose words and deeds are inconsistent because voters react negatively to leaders who break promises. This study examined how much of that punishment stemmed from voters' dislike of broken promises and how much stemmed from voters' dislike of certain actions. In other words, how much is a voter's approval of a leader's policy driven by voter preferences over the consistency between that policy and past commitments and how much is approval driven by the voter's preferences over the policy itself? A survey experiment demonstrated that consistency matters most for citizens without strong policy preferences. For these citizens, audience costs are indeed costly- inconsistency between commitments and policies causes a substantial drop if their approval of leaders. However, consistency has a much smaller effect for citizens who hold stronger policy preferences. For citizens who have opinions supporting or opposing a certain policy, learning of inconsistency in their

[19] Note that prediction metrics based on percentages correctly predicted, such as percent reduction in error, are not applicable here since neither model predicts that any respondents will disapprove.

leader's policy choice does not substantially change their approval of the leader. In other words, citizens with stronger policy opinions do not impose significant audience costs. This is not to suggest that consistency effects are "zero" or irrelevant. Consistency effects were apparent for certain groups, namely respondents without strong policy opinions. But they are significantly moderated for groups with policy opinions.

The finding that audience costs are moderated by preferences over policy has important implications applying ACT to the question of how international institutions and organizations facilitate cooperation. If audience preferences over consistency dominate their preferences over policy, then ACT predicts a robust, consistent effect of international commitments on member state behavior. International agreements and institutions are strong forces for compliance because, once a leader has committed to a certain policy, audiences will react negatively to defections from those obligations, regardless of the audience members preferences over the actual policy. For a leader choosing whether to cooperate with a partner country, their decision calculus in a world where they have committed to cooperate is fundamentally from their decision calculus in a world without that commitment. Irrespective of their domestic constituents' preferences over cooperation, the leader's commitment acts as a strong inducement to choose to honor their commitment by choosing to cooperate.

However, to the degree that preferences over policy endure, even after leaders have made commitments, the effects of those commitments is less pronounced. Consider two "types" of audience members, those who support compliance with international agreements and those who support defection. If ACT is correct and preferences over consistency are strong, then both types of audience members should be equally displeased with leaders who defect from international agreements, regardless of whether they supported compliance with the agreement in the first place. If, on the other hand, policy preferences are strong, then audiences who support cooperation will be more likely to condemn defections and audiences who oppose cooperation will react less negatively (or even positively) to news that their government has broken its obligations. If audience reactions are conditional on audience preferences, then the political calculus of a leader who has not made any previous commitments is very similar to the calculus facing a leader who has made commitments. In both the world with the agreement and the world without that commitment, the leader's decision calculus is largely based on the expressed or anticipated audience preferences over that policy. As preferences over consistency become more important, the effectiveness of international commitments grows unconditionally. As preferences over policy become more important, the effectiveness of commitments becomes increasingly conditioned by the balance of political power between pro- and anti-compliance audiences and the salience of particular issues.

There is likely to be significant variation in the effectiveness of institutions both within and across member states because of variation in preferences over policy. Within member states, institutions and agreements are less effective at changing the opinions of groups with strong policy preferences. For member states with highly polarized domestic groups, some in strong support of compliance with international obligations and some strongly opposed, the presence of an international obligation will have less of an effect on changing public opinion- and in turn, less effect on influencing policymakers beholden to those groups.

This question is likely to be particular important depending on the issue area governed by a particular institution. Some international institutions govern highly salient and polarizing policy areas, such as those dealing with state sovereignty or human rights violations. In the context of human rights abuses or war crimes, domestic audiences are likely to be highly sensitive to the costs and benefits of complying with an international institution that calls for the trial and possible imprisonment of a popular political figure, as is the case with the International Criminal Court. Other institutions govern policy areas which, though important to subsets of the population, are not as salient or important to the population at large. Consider international trade and countries' obligations to refrain from protectionism under the World Trade Organization. Some audiences, such as import-competing producers, might be highly sensitive to compliance these rules. Other audiences, such as consumers who potentially benefit from compliance via lower prices or less deadweight loss, are less sensitive to compliance policy since the benefits are diffuse and small for each individual.

The distinction between preferences over consistency and preferences over policy is even more important in international cooperation than in crisis bargaining, because the two contexts differ in a fundamental way: the ease with which an audience can assess policy choices, and by implication, their consistency with past commitments. In crisis bargaining, the ultimate policy choice is over whether or not to use military force in order to back up a threat. The use of military force is most often a public act- audiences, regardless of their location or level of political sophistication, usually know whether military force has been used or not, and by implication, whether their leader's commitments have been honored.[20] This is in contrast with the context of international cooperation where many issue areas are governed by more opaque policies, and compliance is difficult for audiences to observe. For example, audiences lack information about whether their government's emissions reductions efforts will meet international targets. In international trade, non-tariff barriers are especially inaccessible for the average audience member, with democracies often deliberately obscuring their policies (Kono, 2006). As a result, when audience members learn that their government's policies violate an international agreement, they are not just learning about the consistency between their leader's commitments and actions, but about the actions themselves.

The results from the survey analysis also suggested that the groups most influenced by consistency effects are also most influenced by any other arguments supporting or opposing certain policies. For these groups, even placebo arguments, that contained no argumentative content, were persuasive. This likely dampens the effects of audience costs overall, since audiences are likely to be deluged with pro- and con- arguments for every policy decision of any consequence. Elites in favor of or opposing the policy are always able to find arguments supporting their side's contention, regardless of the validity of those arguments. Levendusky and Horowitz (2012) find that audience costs are significantly lessened when the president claims that his actions were justified by new

[20] To be sure, some military acts are covert. But these cases are already beyond the scope of audience costs theory, since it is anathema for a leader to make a commitment regarding the use of covert military force.

information. In some cases, audiences were *more* supportive of presidents who made a promise, broke it, but justified the decision than they were of presidents who did not make promises. It is highly unlikely that a policymaker would ever break a prior promise or commitment and *not* argue that the decision was justified in some way. If audiences most susceptible to consistency-based arguments are also susceptible to other arguments or *ex post* justifications, then there is no guarantee that consistency-based arguments will win out.

Finally, the results taken together suggest that the challenge for international institutions and agreements is not "How to persuade the malleable?" but rather "How to persuade the intransigent?" An important future task for scholars interested in international cooperation is to determine how international institutions and agreements can persuade domestic audiences who have a strong stake in non-compliance that they should support leaders who enact compliant policies. Institutions need to be more than informational devices that "get the word out." They need to be able to sway stubborn audiences as well as more malleable audiences.

Figure 4.1: Approval by Treatment Group

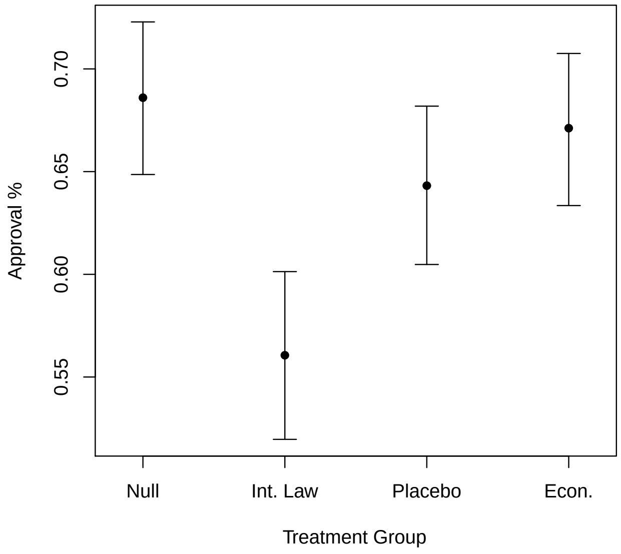


Figure 4.2: Treatment Effects by Trade Preferences: Int. Law vs. Null

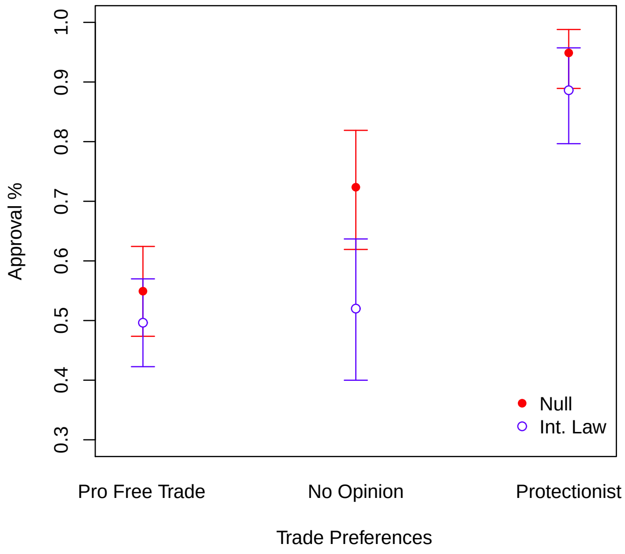


Figure 4.3: Treatment Effects by Trade Preferences: Placebo vs. Null

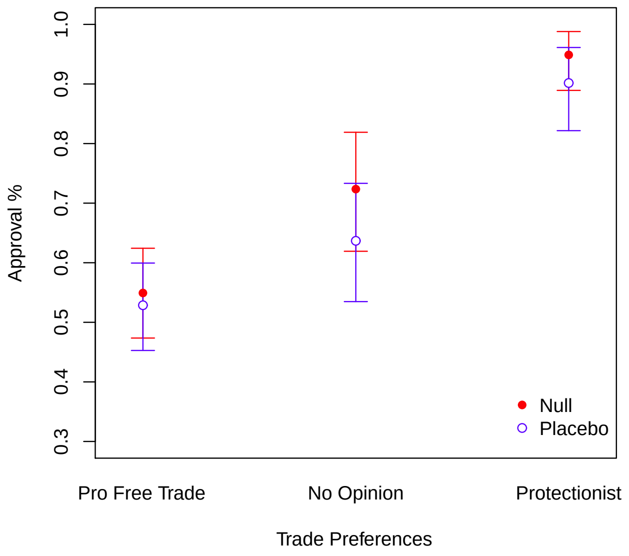


Figure 4.4: Treatment Effects by Trade Preferences: Econ. vs. Null

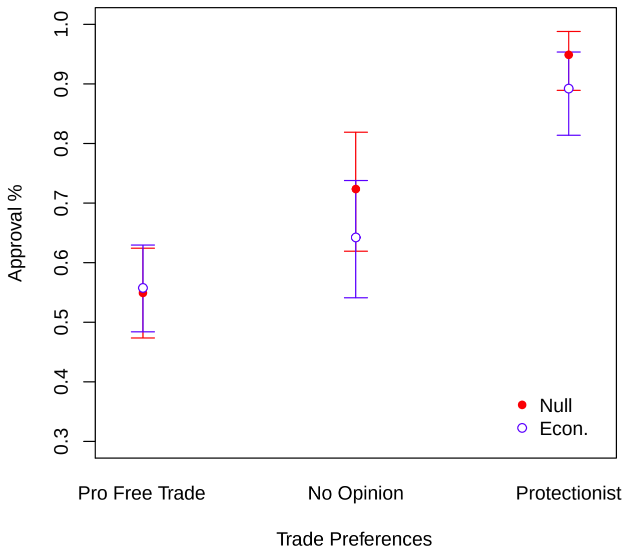


Table 4.1: Effect of All Covariates on Treatment Probability

	Int. Law	Economic	Null	Placebo
Age	-.002	-.004	.008	-.002
	(.006)	(.006)	(.006)	(.006)
Male	.247	-.043	-.098	-.104
	(.120)**	(.118)	(.120)	(.119)
White	-.095	-.219	.182	.176
	(.259)	(.255)	(.283)	(.263)
Black	-.384	-.030	.155	.265
	(.354)	(.329)	(.356)	(.340)
Hispanic	-.195	-.390	-.006	.570
	(.362)	(.362)	(.386)	(.344)*
Asian	-.465	-.342	.752	-.021
	(.482)	(.446)	(.424)*	(.451)
Married	.197	.017	-.172	-.037
	(.138)	(.135)	(.138)	(.137)
College Educ.	-.025	-.120	.076	.072
	(.172)	(.170)	(.177)	(.174)
Polit. Knowledge	.039	-.019	-.077	.063
	(.051)	(.050)	(.051)	(.051)
Isolationism	-.073	.041	-.026	.060
	(.061)	(.061)	(.062)	(.061)
Ethnocentrism	.024	-.020	-.090	.081
	(.068)	(.067)	(.069)	(.067)
Working	-.085	.340	-.027	-.223
	(.118)	(.118)***	(.117)	(.116)*
Above Median Income	-.111	.071	.073	-.020
	(.122)	(.120)	(.121)	(.120)
Republican	.012	.093	-.053	-.008
	(.164)	(.167)	(.167)	(.166)
Conservative	.089	-.236	.060	.109
	(.175)	(.179)	(.179)	(.177)
Pro-taxes	.386	-.011	-.063	-.254
	(.142)***	(.136)	(.135)	(.132)*
Union Member	.063	-.159	.024	.071
	(.118)	(.117)	(.118)	(.116)
Inequality	.021	.169	-.010	-.012
	(.033)	(.131)	(.014)	(.014)
N	1695	1695	1695	1695
χ^2	21.533	19.547	14.519	16.11
p value	.253	.359	.695	.585
pseudo R^2	.011	.01	.008	.008

Table 4.2: Effect of Pre-treatment Covariates on Treatment Probability

	Int. Law	Economic	Null	Placebo
Age	-.0002 (.005)	-.005 (.005)	.005 (.005)	-.0008 (.005)
Male	.201 (.115)*	-.034 (.114)	-.152 (.115)	-.014 (.114)
White	-.079 (.211)	-.113 (.207)	.180 (.224)	.027 (.214)
Black	-.196 (.297)	.011 (.282)	.141 (.301)	.045 (.292)
Hispanic	-.251 (.334)	-.309 (.331)	.040 (.342)	.466 (.310)
Asian	-.494 (.462)	-.278 (.427)	.769 (.391)**	-.122 (.430)
Married	.166 (.130)	.011 (.129)	-.118 (.130)	-.056 (.129)
College Educ.	-.084 (.165)	-.049 (.164)	.103 (.171)	.032 (.167)
N	1752	1752	1752	1752
χ^2	6.817	2.463	7.831	3.891
p value	.557	.963	.45	.867
pseudo R^2	.003	.001	.004	.002

Table 4.3: Effect of Treatment Assignment on Free Trade Responses

International Law Treatment	-.212 (.182)
Economic Treatment	.005 (.174)
Placebo Treatment	.005 (.176)
N	943
χ^2	2.085
p value	.555
pseudo R^2	.001

Table 4.4: Approval Rates by Treatment Group

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	415	0.687				
Int. Law	405	0.560	-0.126	0.034	-3.76	¡0.01
Econ	426	0.671	-0.015	0.032	-0.48	0.633
Placebo	427	0.644	-0.043	0.023	-1.31	0.190

Table 4.5: Approval Rates by Treatment Group and by Respondent Trade Preference

Pro-Free Trade Respondents

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	120	0.550				
Int. Law	123	0.496	-0.054	0.064	-0.84	0.401
Econ	129	0.558	0.008	0.063	0.13	0.898
Placebo	121	0.529	-0.021	0.064	-0.33	0.744

No Opinion Respondents

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	55	0.727				
Int. Law	46	0.522	-0.206	0.095	-2.16	0.033
Econ	62	0.645	-0.082	0.087	-0.95	0.345
Placebo	61	0.639	-0.088	0.087	-1.01	0.314

Protectionist Respondents

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	48	0.958				
Int. Law	39	0.897	-0.061	0.053	-1.11	0.270
Econ	50	0.900	-0.058	0.052	-1.12	0.267
Placebo	45	0.911	-0.047	0.051	-0.92	0.360

Chapter 5

Conclusion

The international community and Western powers in particular have invested a tremendous amount of effort over the past two decades to build and strengthen institutions and agreements designed to facilitate cooperation among nations. This time period has seen the emergence of important institutions like the World Trade Organization's Dispute Settlement Understanding, the continued influence of established bodies like the European Union's Court of Justice, and hope and ambition embodied in the relatively new International Criminal Court. Though institutions like these do not enjoy universal support from every member on every contentious issue, they at least appear to have significant effects on cooperation, often inducing sovereign nations to choose cooperative policies when they might otherwise have been tempted to defect. The zeal with which the international community has pursued such institutions has been based, at least in part, on the theory that these institutions can delineate specific and worthwhile obligations for members and empower domestic audiences to monitor policymakers and convince them to abide by these obligations.

This dissertation has thoroughly explored the promise and limitations of this strategy. Crucially, the ability of institutions to facilitate cooperation is founded upon the preferences and political strength of those domestic audiences. Institutions as alarms serve as *magnifiers* for underlying audience preferences. When politically important audiences support cooperation, institutions which monitor policymakers' decisions can empower these audiences to induce their leaders to choose policies in line with international obligations. But when audiences oppose cooperation or when pro-cooperation audiences are politically irrelevant, institutions that provide monitoring can fail to induce cooperation or even rally certain groups against it.

The challenge moving forward asks: How can institutions become *movers* of preferences, rather than *magnifiers*? Institutions that facilitate monitoring within countries or over issue areas where there is nascent audience support for cooperation have the largest impact in terms of facilitating cooperation when it might not otherwise have occurred. In a sense, these are the "easy cases." They are cases where a little bit of additional information can go a long way towards cooperation.

A tougher challenge lies in the "hard cases," where the problem is not only a lack of information about policymakers' decisions, but also a resistance to cooperation or compliance in the first place. Challenges such as this arise in virtually every area of international cooperation and in many different countries. For example, the challenge in convincing citizens that their leaders should cooperate with the International Criminal Court is not a lack of information. Citizens in Kenya are bombarded with daily news articles about their leaders' efforts to resist complying with the country's ICC obligations. Yet politicians can continue this resistance because they enjoy strong support from constituencies opposed to the ICC process. In the context of global climate change, monitoring countries' compliance with international obligations is a formidable challenge. But a greater challenge lies in convincing citizens and political elites within key countries that the benefits of pollution abatement efforts outweigh the costs.

The bad news for the goal of making institutions movers, rather than magnifiers, of audience preferences is that this task is daunting. Preferences do not arise by accident. Strong economic, political, or ideological interests often underlie citizens' resistance to compliance with international obligations. The good news is that some domestic and international institutions have managed to achieve significant compliance records despite similar challenges. The United States Supreme Court, for example, was not born into an existence where citizens strongly supported compliance with its rulings despite their underlying interests and preferences. It evolved into this role and gained this power over time.

My future research will look at how institutions can successfully become movers of preferences rather than magnifiers. What do existing international relations theories of institutional design, delegation, information, etc. say about how best to achieve this role. What do Americanist and comparativist work on courts and political agency and comparative work on legitimacy and institutional change contribute? What new theories explain the ability or failure of international institutions to move and shape preferences? As the international community seeks to defend the gains made in enhancing cooperation over issues like trade and investment, and seeks to expand into and solidify efforts in new areas like war crimes, human rights, and climate change, these questions are paramount.

Bibliography

Abbott, Kenneth W. and Duncan Snidal. 1998. "Why States Act through Formal International Organizations." *The Journal of Conflict Resolution* 42(1):pp. 3–32.

Alford, Roger P. 2006. "Reflections on US-Zeroing: A Study in Judicial Overreaching by the WTO Appellate Body." *Columbia Journal of Transnational Law* 44.

Allee, Todd L. N.d. "The Hidden Impact of the World Trade Organization on the Reduction of Trade Conflict." Paper for 2005 Midwest Political Association Conference.

Allee, Todd L. and Paul K. Huth. 2006. "Legitimizing Dispute Settlement: International Legal Rulings as Domestic Political Cover." *The American Political Science Review* 100(2):219–234.

Ashworth, Scott and Kristopher W. Ramsay. 2009. "Should Audiences Cost? Optimal Domestic Constraints in International Crises." Working paper.

Bergsten, C. Fred and William R. Cline. 1983. Trade Policy in the 1980's: An Overview. In *Trade Policy in the 1980's*, ed. William R. Cline. The MIT Press.

Berinsky, Adam J., Gregory A. Huber, Gabriel S. Lenz and Edited by R. Michael Alvarez. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* .

Bernheim, B. Douglas and Michael D. Whinston. 1986. "Menu Auctions, Resource Allocation, and Economic Influence." *The Quarterly Journal of Economics* 101(1):1–31.

Blonigen, Bruce A. and Chad P. Bown. 2003. "Antidumping and retaliation threats." *Journal of International Economics* 60(2):249 – 273.

Blonigen, Bruce A. and Thomas J. Prusa. 2001. Antidumping. Working Paper 8398 National Bureau of Economic Research.

Bown, Chad. 2004. "Trade disputes and implementation of protection under GATT: An empirical assessment." *Journal of International Economics* 62(2):263–294.

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

Bown, Chad P. 2005*b*. "Trade Remedies and World Trade Organization Dispute Settlement: Why Are So Few Challenged?" *The Journal of Legal Studies* 34(2):515–555.

Bthe, Tim and Helen V. Milner. 2008. "The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements?" *American Journal of Political Science* 52(4):741–762.

Busch, Marc L. 2007. "Overlapping Institutions, Forum Shopping, and Dispute Settlement in International Trade." *International Organization* 61(04):735–761.

Busch, Marc L. and Eric Reinhardt. 2003. "Developing Countries and General Agreement on Tariffs and Trade/World Trade Organization Dispute Settlement." *Journal of World Trade* 37:719–735.

Busch, Marc L., Eric Reinhardt and Gregory Shaffer. 2009. "Does legal capacity matter? A survey of WTO Members." *World Trade Review* 8(04):559–577.

Canes-Wrone, Brandice and Kenneth W. Shotts. 2004. "The Conditional Nature of Presidential Responsiveness to Public Opinion." *American Journal of Political Science* 48(4):690–706.

Carrubba, Clifford J. 2005. "Courts and Compliance in International Regulatory Regimes." *The Journal of Politics* 67(3):669–689.

Carrubba, Clifford J., Matthew Gabel and Charles Hankla. 2008. "Judicial Behavior under Political Constraints: Evidence from the European Court of Justice." *American Political Science Review* 102(04):435–452.

Carrubba, Clifford James. 2009. "A Model of the Endogenous Development of Judicial Institutions in Federal and International Systems." *The Journal of Politics* 71(01):55–69.

Chang, Eric C. C., Miriam A. Golden and Seth J. Hill. 2010. "Legislative Malfeasance and Political Accountability." *World Politics* 62(02):177–220.

Chapman, Terrence L. 2007. "International Security Institutions, Domestic Politics, and Institutional Legitimacy." *Journal of Conflict Resolution* 51(1):134–166.

Chapman, Terrence L. 2009. "Audience Beliefs and International Organization Legitimacy." *International Organization* 63(04):733–764.

Chapman, Terrence L. and Scott Wolford. 2010. "International Organizations, Strategy, and Crisis Bargaining." *The Journal of Politics* 72(01):227–242.

Chaudoin, Stephen. 2011. "Information Transmission and the Strategic Timing of Trade Disputes." Manuscript, Princeton University.

Dai, Xinyuan. 2002. "Information Systems in Treaty Regimes." *World Politics* 54(4):405–436.

Dai, Xinyuan. 2007. *International Institutions and National Policies*. Cambridge University Press.

Davis, Christina. 2011. "Why Adjudicate? Enforcing Trade Rules." Princeton University.

Davis, Christina L. and Sarah Blodgett Bermeo. 2009. “Who Files? Developing Country Participation in GATT/WTO Adjudication.” *The Journal of Politics* 71(03):1033–1049.

Davis, Christina and Yuki Shirato. 2007. “Firms, Governments, and WTO Adjudication: Japan’s Selection of WTO Disputes.” *World Politics* 59(2):274–284.

Elkins, Zachary, Andrew T. Guzman and Beth A. Simmons. 2006. “Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000.” *International Organization* 60(4):pp. 811–846.

Fang, Songying. 2008. “The Informational Role of International Institutions and Domestic Politics.” *American Journal of Political Science* 52(2):304–321.

Fang, Songying. 2010. “The Strategic Use of International Institutions in Dispute Settlement.” *Quarterly Journal of Political Science* 5(2):107–131.

Fearon, James D. 1994. “Domestic Political Audiences and the Escalation of International Disputes.” *The American Political Science Review* 88(3):577–592.

Gabel, Matthew. 1998. “Public Support for European Integration: An Empirical Test of Five Theories.” *The Journal of Politics* 60(02):333–354.

Gartzke, Erik and Megumi Naoi. 2011. “Multilateralism and Democracy: A Dissent Regarding Keohane, Macedo, and Moravcsik.” *International Organization* 65(03):589–598.

Gawande, Kishore, Pravin Krishna and Marcelo Olarreaga. 2009. “What Governments Maximize and Why: The View from Trade.” *International Organization* 63(03):491–532.

Gilligan, Michael, Leslie Johns and B. Peter Rosendorff. 2010. “Strengthening International Courts and the Early Settlement of Disputes.” *Journal of Conflict Resolution* 54(1):5–38.

Grossman, Gene M. and Elhanan Helpman. 1994. “Protection for Sale.” *The American Economic Review* 84(4):833–850.

Guisinger, Alexandra. 2009. “Determining Trade Policy: Do Voters Hold Politicians Accountable?” *International Organization* 63(03):533–557.

Guzman, Andrew T. and Beth A. Simmons. 2005. “Power Plays and Capacity Constraints: The Selection of Defendants in World Trade Organization Disputes.” *The Journal of Legal Studies* 34(2):pp. 557–598.

Hansen, Wendy L. 1990. “The International Trade Commission and the Politics of Protectionism.” *The American Political Science Review* 84(1):pp. 21–46.

Hays, Jude C., Sean D. Ehrlich and Clint Peinhardt. 2005. “Government Spending and Public Support for Trade in the OECD: An Empirical Test of the Embedded Liberalism Thesis.” *International Organization* 59(02):473–494.

Hiscox, Michael J. 2002. *International Trade and Political Conflict: Commerce, Coalitions, and Mobility*. Princeton University Press.

Horn, Henrik, Petros C. Mavroidis and Hakan Nordstrom. 1999. "Is the Use of the WTO Dispute Settlement Process Biased?" *Center for Economic Policy Research Discussion Paper* 2340.

Imai, Kosuke and David A. van Dyk. 2005. "A Bayesian analysis of the multinomial probit model using marginal data augmentation." *Journal of Econometrics* 124(2):311 – 334.

Irwin, Douglas. 2004. The Rise of U.S. Antidumping Actions in Historical Perspective. Working Paper 10582 National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w10582*

Johns, Leslie. Forthcoming. "Courts as Coordinators: Endogenous Enforcement and Jurisdiction in International Adjudication." *Journal of Conflict Resolution* .

Kelemen, R. Daniel and David Vogel. 2010. "Trading Places: The Role of the United States and the European Union in International Environmental Politics." *Comparative Political Studies* 43(4):427–456.

Keohane, Robert O. 1984. *After hegemony: Cooperation and discord in the world political economy*. Princeton, NJ: Princeton University Press.

Keohane, Robert O., Stephen Macedo and Andrew Moravcsik. 2009. "Democracy-Enhancing Multilateralism." *International Organization* 63(01):1–31.

Keohane, Robert O., Stephen Macedo and Andrew Moravcsik. 2011. "Constitutional Democracy and World Politics: A Response to Gartzke and Naoi." *International Organization* 65(03):599–604.

Kono, Daniel Y. 2006. "Optimal Obfuscation: Democracy and Trade Policy Transparency." *The American Political Science Review* 100(3):369–384.

Koremenos, Barbara. 2012. "Open Covenants, Clandestinely Arrived At." Manuscript.

Krikorian, Jacqueline D. 2005. *Managing the challenges of WTO participation: 45 case studies*. Cambridge University Press chapter Canada and the WTO: Multilevel Governance, Public Policy-Making and the WTO Auto Pact Case, pp. 134–149.

Lam, Patrick. 2007. *coxph: Cox Proportional Hazard Regression for Duration Dependent Variables*. Vol. Zelig: Everyone's Statistical Software.
URL: *http://gking.harvard.edu/zelig*

Leeds, Brett Ashley. 1999. "Domestic Political Institutions, Credible Commitments, and International Cooperation." *American Journal of Political Science* 43(4):pp. 979–1002.

Levendusky, Matthew S. and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74(02):323–338.

Mansfield, Edward D. and Diana C. Mutz. 2009. “Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety.” *International Organization* 63(03):425–457.

Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. 2000. “Free to Trade: Democracies, Autocracies, and International Trade.” *The American Political Science Review* 94(2):305–321.

Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. 2002. “Why Democracies Cooperate More: Electoral Control and International Trade Agreements.” *International Organization* 56(3):477–513.

Mansfield, Edward D. and Marc L. Busch. 1995. “The political economy of nontariff barriers: a cross-national analysis.” *International Organization* 49(04):723–749.

Margalit, Yotam. 2011. “Costly Jobs: Trade-related Layoffs, Government Compensation, and Voting in U.S. Elections.” *American Political Science Review* 105(01):166–188.

McCubbins, Mathew D., Roger G. Noll and Barry R. Weingast. 1987. “Administrative Procedures as Instruments of Political Control.” *Journal of Law, Economics, & Organization* 3(2):pp. 243–277.

Milgrom, Paul R., Douglass C. North and Barry R. Weingast. 1990. “The Role of Institutions in the Rivival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs.” *Economics & Politics* 2(1):1–23.

Milner, Helen V. and Dustin H. Tingley. 2011. “Who Supports Global Economic Engagement? The Sources of Preferences in American Foreign Economic Policy.” *International Organization* 65(01):37–68.

Milner, Helen V. and Keiko Kubota. 2005. “Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries.” *International Organization* 59(1):pp. 107–143.

Nordhaus, William D. 1975. “The Political Business Cycle.” *The Review of Economic Studies* 42(2):pp. 169–190.

Peer, Eyal, Gabriele Paolacci, Jesse Chandler and Pam Mueller. 2012. “Screening participants from previous studies on Amazon Mechanical Turk and Qualtrics.” Unpublished Manuscript.

Prusa, Thomas J. 1992. “Why are so many antidumping petitions withdrawn?” *Journal of International Economics* 33(1-2):1 – 20.

Rickard, Stephanie. 2010. “Democratic Differences: Electoral Institutions and Compliance with GATT/WTO Agreements.” *European Journal of International Relations* 16(4):711–730.

Rogowski, Ronald. 1987. “Political Cleavages and Changing Exposure to Trade.” *The American Political Science Review* 81(4):pp. 1121–1137.

Rose, Andrew K. 2004. “Do We Really Know That the WTO Increases Trade?” *The American Economic Review* 94(1):98–114.

Rosendorff, B. P. 2005. “Stability and Rigidity: Politics and Design of the WTO’s Dispute Settlement Procedure.” *American Political Science Review* 99(03):389–400.

Sabry, Faten. 2000. “An Analysis of the Decision to File, the Dumping Estimates, and the Outcome of Antidumping Petitions.” *The International Trade Journal* 14(2):109–145.

Sattler, Thomas and Thomas Bernauer. 2011. “Gravitation or discrimination? Determinants of litigation in the World Trade Organisation.” *European Journal of Political Research* 50(2):143–167.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy* . New York: Cambridge University Press.

Simmons, Beth A. 2000. “International Law and State Behavior: Commitment and Compliance in International Monetary Affairs.” *The American Political Science Review* 94(4):819–835.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.

Simmons, Beth A. 2010. “Treaty Compliance and Violation.” *Annual Review of Political Science* 13:273–296.

Simmons, Beth A. and Allison Danner. 2010. “Credible Commitments and the International Criminal Court.” *International Organization* 64(02):225–256.

Slantchev, Branislav. 2006. “Politicians, the media, and domestic audience costs.” *International Studies Quarterly* 50:445–477.

Smith, Alastair. 1998. “International Crises and Domestic Politics.” *The American Political Science Review* 92(3):pp. 623–638.

Snyder, Jack and Erica D. Borghard. 2011. “The Cost of Empty Threats: A Penny, Not a Pound.” *American Political Science Review* 105(03):437–456.

Staton, Jeffrey K. 2006. “Constitutional Review and the Selective Promotion of Case Results.” *American Journal of Political Science* 50(1):98–112.

Stone, Alec. 1992. *The Birth of Judicial Politics in France*. Oxford: Oxford University Press.

Sueyoshi, Glenn T. 1992. “Semiparametric proportional hazards estimation of competing risks models with time-varying covariates.” *Journal of Econometrics* 51(1-2):25 – 58.

Tomz, Michael. 2007. “Domestic Audience Costs in International Relations: An Experimental Approach.” *International Organization* 61(04):821–840.

Tomz, Michael. 2008. “Reputation and the Effect of International Law on Preferences and Beliefs.” Stanford University.

Tomz, Michael and Robert P. Van Houweling. 2012. “Candidate Repositioning.” Unpublished Manuscript.

Tussie, Diana and Valentina Delich. 2005. *Managing the challenges of WTO participation: 45 case studies*. Cambridge University Press chapter Dispute Settlement between Developing Countries: Argentina and Chilean Price Bands, pp. 23–37.

Vanberg, Georg. 1998. “Abstract Judicial Review, Legislative Bargaining, and Policy Compromise.” *Journal of Theoretical Politics* 10(3):299–326.

Vanberg, Georg. 2001. “Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review.” *American Journal of Political Science* 45(2):pp. 346–361.

Vanberg, Georg. 2005. *The Politics of Constitutional Review in Germany*. Cambridge University Press.

Vandenbussche, Hylke and Maurizio Zanardi. 2010. “The chilling trade effects of antidumping proliferation.” *European Economic Review* 54(6):760 – 777.

Weeks, Jessica L. 2008. “Autocratic Audience Costs: Regime Type and Signaling Resolve.” *International Organization* 62(01):35–64.