Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions

Stephen Chaudoin*

University of Pittsburgh

Word Count: 12,325

August 30, 2012

*I appreciate the helpful comments and suggestions from Michael Aklin, Phil Arena, Terrence Chapman, Christina Davis, Rex Douglass, Michael Horowitz, Sarah Hummel, Matthew Incantalupo, Robert Keohane, Jeffrey Kucik, John Londregan, Helen Milner, Emily Ritter, and Tom Scherer. I am also grateful to The Niehaus Center for Global Governance which provided support for this research.

Abstract

Is a policymaker's audience more concerned with the consistency between words and deeds or with policy itself? A key assumption of audience costs theories of crisis bargaining and international cooperation is that audience members have strong preferences for consistency between their leader's commitments and their leader's actual policy choices. However, audience members are also likely to have strong preferences over the policy choices in and of themselves, regardless of their consistency with past commitments. I conduct a randomized survey experiment to evaluate the magnitude of *consistency* and *policy* effects in the context of international agreements over trade policy. Respondents with expressed policy preferences, whether supporting or opposing free trade, have muted reactions to learning that their leader has broken an agreement. Only respondents with no opinion on trade policy are affected by learning that their leader's policy is inconsistent with prior commitments. This suggests that the ability of audience costs and preferences for consistency to affect policymaker interests is constrained by the underlying preferences of their constituents.

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 international cooperation and crisis bargaining. In the former context, policymakers commit to certain policies when they negotiate, sign, and ratify international agreements or join an international institution. ACT argues that audiences punish policymakers who choose non-compliant 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.[1]

While the term audience costs has expanded to encompass many meanings, a key, common 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, James Fearon (1994) argues that inconsistency creates the opportunity for domestic political opponents to criticize the incumbent for damaging the country's international "credibility, face or honor" (581).[2] In the context of international law and cooperation, 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.[3]

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

[1] Leeds (1999); Mansfield and Pevehouse (2006).

[2] 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. For a recent survey of audience costs arguments, see Snyder and Borghard (2011).

[3] Abbott and Snidal (1998, 428).

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 international cooperation and crisis bargaining 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- a negative reaction to policies that diverge from past promises- and a policy effect- a negative reaction to divergence from the audience's preferred policy. Both consistency and policy effects are likely to affect audience reactions, but decomposing and understanding the relative magnitude of the two 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 theorized by ACT, then this is a cause for optimism about the effects of agreements: audiences, because of their penchant for consistency, are powerful forces for compliance with agreements. However, to the degree that policy effects are important for audience reactions, then the effects of international institutions on member state policy are more 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 more circumscribed.

To distinguish between consistency and policy effects, I embedded an experiment in a survey conducted in May and June 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 (pros) and opposed to (cons) 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. 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 policy preferences. I also present a follow up survey where similar results obtain using primed, rather than elicited, trade 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. To further examine the sources of public support for international commitments I conducted a follow-up survey experiment where respondents were treated with different rationales for opposing a treaty-violating policy. The possibility of foreign retaliation and an aversion to breaking the rule of law have stronger effects on public opinion than do an aversion to inconsistency.

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, through their potential to activate audiences who prefer consistency, 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.

Preferences Over Consistency and Policy

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.[4] Audience costs have alternatively been described as "the surge in disapproval that would occur if a leader made commitments and did not follow through,"[5] 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."[6] In

[4]Fearon (1994).
[5]Tomz (2007, pp. 823).
[6]Levendusky and Horowitz (2012, pp. 324).

democracies, the punishment is often thought of as electoral: voters are less likely to return promise-breaking leaders to office, though audience costs have been analyzed in autocratic settings as well.[7] 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. In the context, of international cooperation, 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.[8]

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 promises. 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.[9] These policy adjustments impact audience members differently, creating winners who benefit from

[7] Weeks (2008).
[8] For a more extensive review of this argument, see Simmons (2010).
[9] Keohane (1984).

the policy adjustments and losers who do not.

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, agreements governing trade policy adjustments have distributional impacts: raising and lowering tariffs, increasing or decreasing subsidies, or changing monetary policy benefits some audience members at the expense of others.[10] The perceived or actual effects of trade policy adjustments have been linked to support or opposition to policies and candidates, legislative voting patterns, and the political cleavages that arise regarding trade policy.[11] A rich body of literature examines variation in support for European integration both across and within countries.[12] A similar body of literature examines variation in domestic political support for international cooperation on climate change and the environment.[13]

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.

This dynamic is demonstrated clearly in the recent example of Kenya's experience with the International Criminal Court (ICC). In 2005, the government of Kenya ratified the Rome Statute, which exposed Kenyan nationals to prosecution by the ICC should they commit war crimes, crimes against humanity, or genocide. During and after the 2007 presidential

[10]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.

[11]Rogowski (1987); Hiscox (2002); Milner and Tingley (2011); Margalit (2011).

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

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

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?" Unsurprisingly, support for the ICC was largely driven by the regional variation in underlying support for particular politicians who had been indicted. In regions where indicted politicians enjoy significant public support, the public was 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 received stronger support. 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.[14] It is highly likely that this variation is driven by preferences over policy, i.e. whether to support a court that antagonized liked or dislike politicians, not preferences over consistency, i.e. whether to support the court because of Kenya's past promises. 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.[15]

Even in the canonical ACT context, crisis bargaining, audiences have strong policy preferences. Audiences care about the policy 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.

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

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

It signals the possibility of military action, however remote, and is an inherently coercive approach to foreign policy. A 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). This characterization suggests that policy preferences condition the degree of audience punishment for inconsistency.

Existing Micro-level Evidence of Consistency Effects

The two most well-known empirical studies of the micro-foundations of audience costs, from Tomz (2007); Levendusky and Horowitz (2012), were in the context of crisis bargaining and were designed to detect consistency effects but not policy effects. In both studies, survey participants were told about an international crisis where one foreign country, the aggressor, considered invading its neighbor country. In the treatment group, participants were 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 were told that their leader's words and deeds were inconsistent. Participants assigned to the control group were told that the aggressor considered invading its neighbor, but the United States' leader elected to stay out of the crisis- implicitly neither threatening nor using military action- and the aggressor proceeded with the invasion. All participants

were then asked whether they approved or disapproved of the president's actions. As predicted by audience cost theory, approval was significantly lower in the treatment group.[16] In Tomz (2007), respondents 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 survey design. Respondents learn that the president is guilty of commitment-policy inconsistency in the treatment scenario, but not in the control scenario, which affects 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

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

creates the possibility that disapproval stems from the respondent's dislike of inconsistency, dislike of policy, or both.

Two additional studies use approaches more closely resembling the one used in this paper. Tomz (2008) analyzes a survey where respondents were randomly assigned to treatments which consisted of different arguments for or against an embargo on imports from Burma. Respondents who were given an argument against the embargo- that it violated international law- were 17% more likely to oppose the embargo than respondents who did not receive this argument. In related American politics research, Tomz and Van Houweling (2012) examine how voters respond to candidates who reposition, i.e. change their stance on an issue. They conduct survey research analyzing *valence* and *proximity* effects of candidate repositioning on voter opinions. Similarly to the effect posited by ACT, candidate repositioning negatively affects voters perceptions of the candidates along a valence dimension. Repositioning might also bring the candidate closer to or further away from the voter's most preferred policy, i.e. a proximity effect. Tomz and Van Houweling (2012) use experiments where respondents read about candidates' positions on taxes and abortion over time. 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.

Hypotheses and Experimental Design

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? To answer this question, I embedded a randomized experiment in a survey conducted in May of 2012. The basic structure of the survey was as follows. First, respondents were

randomly assigned treatment consisting of certain pros and cons of a particular policy, with one con arguing that the policy was inconsistent with past promises. Second, respondents were told that their politician enacted that policy and were asked whether they approved or disapproved of the politician's decision. Third, respondents were asked a lengthy set of demographic and opinion questions. Embedded in this set was a question that more directly elicited the respondent's preferences over the particular policy from the initial experiment.[17]

The goal of the experiment is to assess the relationship between two effects: (1) a consistency effect, whereby respondents express lower approval of a policy when they learn of its inconsistency with past promises and (2) a policy effect, whereby respondents' approval of a policy is governed by their preferences over the policy itself, regardless of consistency. I am interested in two questions. First, what is the relative magnitude of the two effects? As described above, understanding whether preferences over consistency can overcome preferences over policy is important for assessing the possibility that international commitments can help influence leader behavior beyond the underlying preferences of domestic constituents. Second, does the strength of a respondent's policy preferences moderate consistency effects? Respondents with stronger policy preferences should be less affected by treatments pertaining to inconsistency. For respondents who already oppose a policy, learning of that policy's inconsistency with past promises should have minimal effect. Treatments pertaining to inconsistency simply provide "yet another" reason to oppose the policy and just move the respondent marginally closer to a floor level of approval. Similarly, for respondents who strongly support a policy, an inconsistency treatment has to "pull against" the respondent's underlying preferences. Learning about inconsistency should have the strongest effect for respondents without strong preferences over the underlying policy.

[17] As described in more detail later, I also ran a follow-up survey in which respondent preferences were primed, rather than elicited.

Survey Recruitment

Approximately 3,500 survey respondents were recruited using Amazon's Mechanical Turk (mTurk) service. mTurk provides access to a recruitment pool for survey respondents by promising compensation for completion of a particular task- in this case, taking a survey.[18] The advantages of using mTurk are that it is very efficient way of administering surveys without sacrificing much in terms of representativeness. 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.[19] In this study, the respondent pool was relatively close to nationally representative surveys, though, unsurprisingly for an internet-based survey, respondents tended to be younger (average age for this survey was 31.9 years, compared to 49.7 in the 2008 American National Election Survey- ANES), more likely to be male (44.9% male, compared to 42.8% male in the 2008 ANES), and more likely to never have been married (35.2% compared to 14.2% in the 2008 ANES).

Main Experiment

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

[18] In this survey, compensation ranged from \$0.55 to \$0.75.

[19] 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.

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. I specified the European continent to avoid the risk that responses were influenced by the respondent's perceptions of the United States' more politically charged import partners, like China. Respondents were then told that the president had to decide whether to impose a 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 cons- arguments opposing import restrictions- or a null treatment- the respondent was not given a con. The text of the three cons is given below:

- **International Agreement 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 agreement (IA) 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. To be sure, respondents could react to the IA treatment for a variety of reasons, even beyond a dislike of inconsistency. In a follow-up survey described below, I explore in greater detail *why* respondents care about violations of international agreements.

The argument in favor of import restrictions most commonly invoked by politicians is that the restrictions will help save domestic 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 IA 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 they "leaned towards" approving or disapproving. This measure of approval closely resembles that of Tomz (2007) and Levendusky and Horowitz (2012).[20] 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 effect of the IA treatment measures consistency effects. 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 agreement consistency verses economic costs), 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)?

[20]The only differences are 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.

Other Questions and Survey Balance

I also asked a series of questions before and after the main experiment. Before the main experiment, I asked the respondents their age, sex, marital status, and state of residence. After the main experiment, respondents answered a series of opinion questions and demographic questions. Embedded in the post-experiment series was a a question pertaining directly to trade policy, as described in greater detail below.

In the opinion and demographic section, respondents first answered a series of five political knowledge questions that measured their familiarity with certain world events.[21] I then asked a series of questions designed to measure the respondent's preferences over isolationism.[22] 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.[23] I also asked a series of standard demographic questions, such as the respondent's party, ideology, income, education, etc. Finally, 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, whether the respondents were currently employed, and whether they or a family member had ever been a member of a trade union.[24]

I first checked that treatment assignment was not skewed among the covariates measured from these pre- and post-experiment questions. For each of the four treatment groups, I

[21]These were factual, multiple-choice questions with one correct answer. 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)?

[22]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."

[23]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.

[24]Mansfield and Mutz (2009).

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 1.

For each treatment group, I cannot reject the null hypothesis that the coefficients are jointly 0. The χ^2 statistics and associated p-values for each treatment group are: International Agreement- 17.00 (p=0.385), Economic- 22.53 (p=0.127), Placebo- 17.28 (p=0.367), and Null- 11.00 (0.809). 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, as displayed in Table 2. The χ^2 statistics and p values are even lower in these regressions: International Agreement- 6.30 (p=0.614), Economic- 4.79 (p=0.780), Placebo- 6.16 (p=0.630), and Null- 6.93 (p=0.544). The only pre-treatment covariate that was significant was that slightly more Asian respondents received the Null treatment.

To check that respondents actually received the desired treatment, at the very end of the survey, 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. 85.9% were able to correctly recall that they had been given a pro-import restriction argument pertaining to layoffs by the U.S. firm, among a list containing the correct answer and two fabricated arguments in favor of import restrictions. 62.1% 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, unsurprisingly, was the weakest, with only 49.3% of respondents correctly recalling it. The IA and economic treatments were stronger, with 66.7% and 68.5% correctly recalling the con arguments that they'd been given. 63.8% 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 0.33 (for the pros) or 0.25 (for the cons), at the 0.01 level.[25]

Trade Policy Preferences

To measure policy preferences, I also asked a standard free-trade question in the middle of the post-experiment questions. 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."[26] 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.

Eliciting respondents' preferences over free trade allows me to compare the relative magnitudes of consistency effects and policy preference effects. Is respondent approval driven more by the treatment they receive or their underlying preferences over import restrictions? I can also compare the magnitude of treatment effects across respondents with different policy preferences. Is the effect of the IA treatment the same for pro-free trade and protectionist respondents as for no opinion respondents? To the degree that policy preferences moderate

[25]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.

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

consistency effects, the effect of the IA treatment should differ 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. In the absence of the IA treatment, some unknown factors underlie these respondents' support for import restrictions. The IA treatment must overcome these factors to move these respondents to disapprove of import restrictions. Respondents who strongly oppose import restrictions should also show weakened IA treatment effects. For various reasons, these respondents already have a low approval level of import restrictions, so the IA treatment is just another reinforcement of their existing opinions. Respondents with strong preferences supporting or opposing import restrictions 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.

Since the trade policy 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 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 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. 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 1.38 with an associated p value of 0.71.[27] These (null) results are robust to alternate specifications, like multinomial regressions or difference in means tests for trade policy responses across treatment groups.

Experimental Results

Before 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 1 and Table 4 show the percentage of respondents who approved of presidents who implemented import restrictions across each of the treatment groups.[28] Among those who received the null treatment, 68.8% approved of the president's actions. Among those receiving the IA treatment, only 58.0% approved of the president's actions. The difference between the null approval rates and the IA 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 10.8% lower in the IA treatment group than in the null group. This difference is highly

[27]Originally, I randomly selected half of the respondents to receive this question. In case the 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 results from 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 in order to have more respondents of each policy preference type in each treatment category.

[28]All figures show Bayesian estimates of the posterior distribution of the proportion of respondents approving of the president's actions. Let θ_t be the proportion of respondents approving of the president's actions under treatment regime $t \in \{$ Null, International Agreement, Econ, Placebo $\}$. Let n_t be the number of respondents receiving treatment t and a_t be the number of respondents in regime t approving. For a prior distribution for θ_t, I use the non-informative Jeffrey's prior, $\theta_t^0 \sim \beta(0.5, 0.5)$. The conjugate posterior distribution for θ_t is $\theta_t^p \sim \beta(a_t + 0.5, n_t - a_t + 0.5)$. The mean and 95% credibility intervals are from 5,000 draws from the posterior distributions.

statistically significant (p value for the difference in proportion approving is < 0.01).[29][30]

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 67.5%. Even direct economic concerns, like the possibility of job loss in other industries, does not appear to influence approval rates. Among those who received the placebo treatment, 64.5% approved, a 4.1% drop compared to the null group. Neither of these differences is significant at conventional levels.

These initial results appear to be a strong reconfirmation of consistency effects. 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 2 shows the approval rates for the IA 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 IA, economic and placebo treatments are shown numerically in Table 5.

For pro-free trade respondents and protectionist (anti-free trade) respondents, the difference between approval rates for the null group and the IA group are small and insignificant. Among pro-free trade respondents, the approval rates in the IA treatment group were 52.4% compared to 57.6% for the null treatment group. The difference, -5.2% is approximately half as large as the difference found for the entire sample (-10.8%), and is statistically in-

[29]The standard deviation, t stat, and p values for differences in approval rates use the normal approximation of the Bernoulli data. The number of respondents in each group is much larger than traditional minimum values for use of the normal approximation.

[30]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.

significant (p value = 0.309). Among protectionist respondents, the approval rates for the IA treatment group were 88.9% compared to 95.2 for the null treatment group. This difference is also approximately half as large as found in the full sample and is statistically insignificant (p value = 0.211).

The treatment effect found in the full sample is strongly driven by respondents with no preferences over free trade. Among respondents who neither supported nor opposed free trade, the approval rates in the IA group were 59.5%, compared to 73.5% for the null group. The difference of −14.0% is substantively large and statistically significant (p value = 0.059).

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 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 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 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, 1%. It has a larger and negative effect among no-opinion and protectionist respondents, -6.1% and -7.1% respectively, though both are insignificant.

Among pro-free trade respondents, the difference between approval rates in the placebo and null groups was very small and insignificant. Yet for respondents expressing no opinion, the placebo treatment managed to decrease approval by -7.6%, though this difference falls short of conventional significance levels (p value = 0.288).[31]

The strength of the placebo treatment for respondents without strong policy opinions suggests that the effect of the IA 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 IA treatment.[32] In other words, limiting our analysis only to the respondents where we found a significant IA treatment effect, the IA treatment effect was statistically indistinguishable from the placebo treatment effect. In his 2008 study, Michael 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 IA and placebo treatments have an addition effect relative to the null treatment, since the respondent

[31]Interestingly, protectionist respondents were also influenced by the placebo treatment, which decreased approval relative to the null by -8.6%, with an associated p value of 0.095.

[32]The same could be said of the economic treatment effects.

receives no con arguments with the null treatment. Substitution effects arise when comparing approval rates, substituting one argument for another. For no opinion respondents, approval rates are only -6.3% lower under the IA treatment than under the placebo treatment, and this difference is statistically insignificant (p value = 0.407). While we can confidently say that both the IA and placebo treatments have additive effects, we can less confidently say that the IA 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 OLS, regressing the respondent's approval on dummies indicating which treatment the respondent received yields a very small R^2 value of 0.0061. Regressing approval on the respondent's expressed preferences over free trade, however, yields an R^2 value 0.0684, increasing the explained variation in approval by a factor of approximately 11. Logit regressions yield similar pseudo-R^2 values of 0.0615 and 0.0057 for policy effects and consistency effects respectively. The AIC and BIC are much lower for the logit policy effects regression, 1653.095 and 1668.725, than for the consistency regression, 2754.038 and 2776.69.[33]

Follow Up Survey: Primed vs. Elicited Preferences

In the above analyses, I first conducted the main experiment and then elicited respondents' preferences over free trade policy. I then checked whether the treatment administered in the experiment "contaminated" respondents' elicited preferences, and did not find any evidence of these effects. To further ensure that the treatment administered did not affect

[33]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.

respondents' expressed preferences, I conducted a follow-up experiment in July of 2012 in which respondents were *primed* with anti-free trade preferences, rather than asked about their preferences. I randomly assigned half of the respondents to the "primed" group and half to the "un-primed" group. Respondents in the unprimed group took the same survey as above, with random assignment to only the IA and null treatments. Respondents in the primed group also took the same survey, but before reading the vignette about tariffs and being randomly assigned to the IA or null treatment, they answered a series of questions that primed them to dislike free trade. Specifically, they first answered whether they were employed or unemployed and whether they had been unemployed at any time over the past five years. I then asked them to estimate the U.S. unemployment rate as in the main analysis, with the additional prime that "As you may know, the U.S. economy has performed poorly over the last few years." Finally, I asked an intentionally-loaded question that linked employment with trade policy: "As you may know, international trade has increased substantially in recent years. Some people argue that increased international trade causes some U.S. workers to lose their jobs because of increased competition from cheap foreign labor. Do you think it is best to... (A) Raise barriers to trade in order keep U.S. workers from losing their jobs in the first place. (B) Provide additional assistance to those workers to find new jobs. (C) Ensure that the U.S. doesn't make any international commitments which limit our flexibility in dealing with these issues. or (D) All of the above."[34]

The follow-up survey allows me to compare the effects of the IA treatment by whether the respondent was primed or unprimed. I expect that the IA treatment will have a weaker effect on primed respondents, since the priming questions heighten the weight that the respondent places on policy as opposed to consistency. Unprimed respondents should behave similarly to those analyzed in the main experiment above.

[34]This also provides a built-in manipulation check that the priming is indeed influencing respondents' opinions on free trade. 40% answered "All of the above" to this question, which is high considering that only approximately 18% expressed anti-free trade preferences in the main analysis.

The results are very consistent with this prediction and are displayed in Figure 5 and Table 6.[35] For the unprimed group, the IA treatment has an almost identical treatment effect as before. The approval rate for unprimed respondents who received the null treatment was 70.2% compared with 60.3% for unprimed respondents who received the IA treatment. The treatment effect for unprimed respondents was thus approximately -10% which is almost identical to the effect found in the main analysis. For primed respondents, the null approval rate was 65.8% compared with 63.4% for primed respondents who received the IA treatment, a difference of only -2.4%. In other words, the priming questions both decreased the null approval rate and substantially dampened the effect of the IA treatment.

Follow Up Survey: Why Support International Agreements?

So far, this analysis has built on existing work assuming that a preference for consistency was a key reason that audiences opposed the breaking of international agreements. The international agreement treatment used in the main experiment was designed to tap into this concept. Respondents did indeed express lower levels of approval for policymakers whose actions were inconsistent with international agreements.

This treatment, however, could have also tapped into other reasons why respondents support international agreements, apart from their penchant for consistency. For example, respondents could also disapprove of breaking international agreements because they fear retaliation from other members of the agreement. If the respondent thinks that they or their community or country could be harmed by foreign punishment resulting from the breaking of an agreement, then the respondent could disapprove regardless of their desire for consistency. In 2002, orange growers and textile manufacturers in the United States were acutely aware that they were likely targets should the European Union decide to retaliate against U.S. steel tariffs. Similarly, some respondents might simply support the rule of law and dislike

[35] Confidence intervals and statistical tests conducted identically to those in the main experiment.

any action that is perceived to be illegal. None of these reasons for disapproval- consistency, retaliation, or rule of law- are mutually exclusive. Audiences might disapprove of breaking international agreements for any subset of those three reasons.

To analyze which of these three factors most influenced respondent approval, I conducted a follow-up survey experiment of approximately 500 respondents in July of 2012. The experiment was conducted in the exact same way as the main experiment above, except it employed three treatments that were specific to particular reasons why a respondent might disapprove of breaking international agreements. Each respondent was randomly assigned to one of three treatments pertaining to international agreements or a null treatment, as above. The three international agreement treatments each began with "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..." They differed by the reason given for disapproving of breaking the international agreement. The three specific international agreement treatments were:

- **Consistency**: ... As a result, the restrictions would break a promise made to Europe, and we would be going back on our word.
- **Retaliation**: ... As a result, Europe would retaliate by imposing restrictions against U.S. products, which would hurt the U.S. economy.
- **Legality**: ... As a result, the court at the World Trade Organization would rule that these restrictions violate international law.

As in the main experiment, all three were very similar in word count, sentence structure, and the forcefulness of language used. By comparing approval levels for each of the three treatments against the null treatment, I can asses the relative treatment effects of each as reasons for disapproving of breaking international agreements.

Which of the three treatments affected respondents' approval levels? In short, all three, though retaliation and legality had slightly stronger and more significant effects than consistency, as shown in Figure 6 and Table 7.[36] All three treatments lowered approval relative to the null treatment by $10-12\%$. The difference in mean approval levels between the retaliation and legality treatments and the null treatment were statistically significant, though the difference between the consistency and null treatments just missed conventional significance (p value $= 0.109$).[37]

These results indicate that respondents' reasons for disapproving of violations of international agreements are likely to be multifaceted, not simply based on a dislike of inconsistency. Respondents were most influenced by the possibility of foreign retaliation, which is a cooperation-facilitating mechanism that, ironically, does not require an international agreement. Countries can use the threat of punishment and retaliation as inducements for cooperation even outside of the purview of international law or agreements.

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

[36] Confidence intervals and statistical tests conducted identically to those in the main experiment.
[37] The treatment effects were also very similar to those found in the main experiment.

between commitments and policies causes a substantial drop in their approval of leaders. However, consistency has a much smaller effect for citizens who hold stronger policy preferences. For citizens with opinions supporting or opposing a certain policy, learning of inconsistency in their 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 for applying ACT to the question of how international institutions and organizations facilitate cooperation. To the degree that audience preferences over consistency can overcome 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 promise 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 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.[38] 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

[38]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.

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 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.

References

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

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

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* .

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.

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

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.

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

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. and Jon C. Pevehouse. 2006. "Democratization and International Organizations." *International Organization* 60(01):137–167.

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

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.

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

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

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

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

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.

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.

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

Figure 1: Treatment Effects, All Respondents

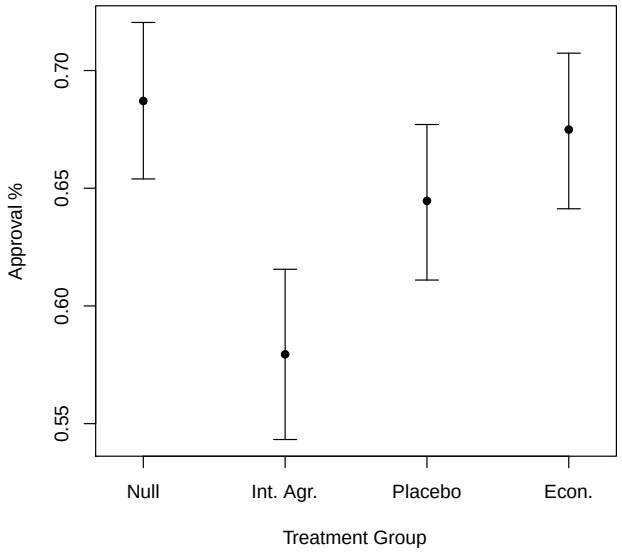

Figure 2: International Agreement Treatment Effects, by Respondent Trade Preferences

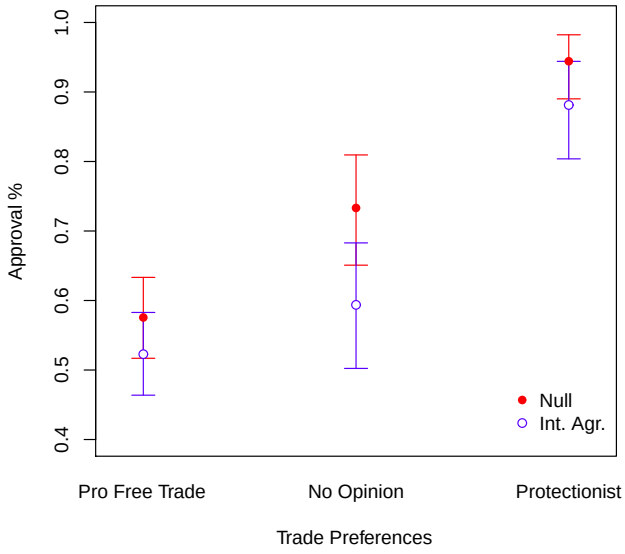

Figure 3: Placebo Treatment Effects, by Respondent Trade Preferences

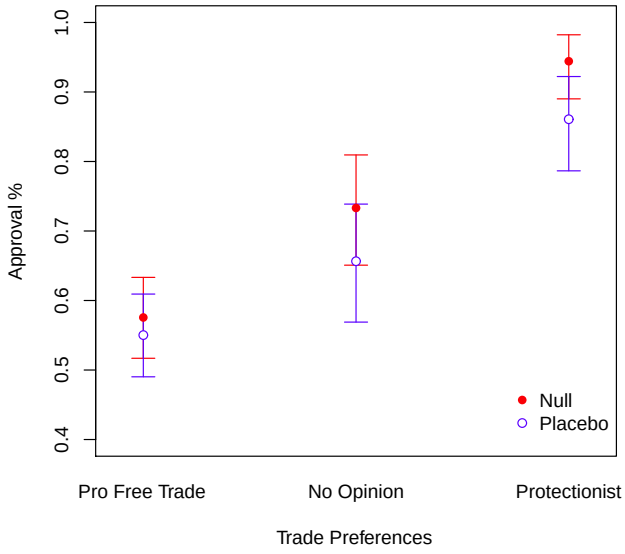

Figure 4: Economic Treatment Effects, by Respondent Trade Preferences

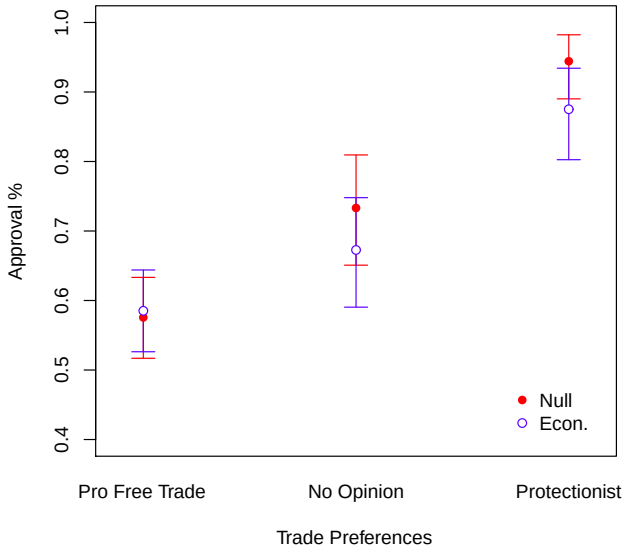

Figure 5: International Agreement Treatment Effects, Unprimed vs. Primed Respondents

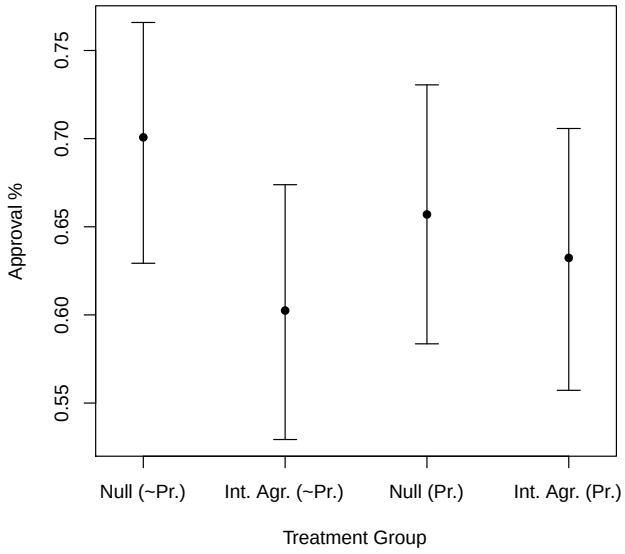

Figure 6: Specific International Agreement Treatment Effects, All Respondents

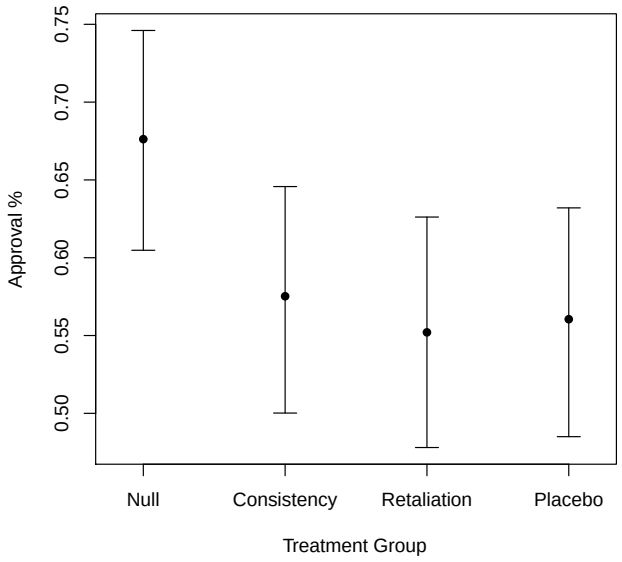

Table 1: Effect of All Covariates on Treatment Probability

	Int. Agr.	Economic	Null	Placebo
	(1)	(2)	(3)	(4)
Age	-.002	-.004	.007	-.0008
	(.005)	(.005)	(.005)	(.005)
Male	.176	.012	.025	-.099
	(.106)*	(.036)	(.056)	(.104)
White	.036	-.297	.032	.225
	(.227)	(.218)	(.233)	(.226)
Black	-.084	-.299	.053	.330
	(.308)	(.293)	(.307)	(.297)
Hispanic	.041	-.560	-.074	.546
	(.308)	(.312)*	(.319)	(.296)*
Asian	-.392	-.647	.650	.247
	(.420)	(.402)	(.356)*	(.373)
Married	.245	-.001	-.198	-.031
	(.124)**	(.122)	(.125)	(.124)
College Educ.	.054	-.166	.005	.113
	(.156)	(.150)	(.156)	(.156)
Pol. Know. Sum	-.005	-.041	-.030	.070
	(.045)	(.044)	(.045)	(.045)
Isolationism	-.092	.042	-.028	.074
	(.054)*	(.053)	(.054)	(.053)
Ethnocentrism	.011	-.062	-.067	.109
	(.060)	(.060)	(.061)	(.059)*
Working	-.129	.300	-.048	-.129
	(.104)	(.104)***	(.104)	(.103)
Above Med. Income	-.177	.116	.101	-.040
	(.108)*	(.105)	(.107)	(.106)
Republican	.111	.039	.032	-.195
	(.144)	(.143)	(.146)	(.147)
Conservative	.015	-.217	.075	.128
	(.156)	(.158)	(.158)	(.159)
Union	.106	-.189	.031	.051
	(.104)	(.104)*	(.104)	(.103)
N	2156	2156	2156	2156
chi^2	17.006	22.531	11.006	17.284
p value	.385	.127	.809	.367
pseudo R^2	.007	.009	.005	.007

Table 2: Effect of Pre-treatment Covariates on Treatment Probability

	Int. Agr.	Economic	Null	Placebo
	(1)	(2)	(3)	(4)
Age	.0009	-.005	.005	-.0006
	(.005)	(.005)	(.005)	(.005)
Male	.131	.012	.021	-.051
	(.102)	(.034)	(.046)	(.060)
White	.043	-.111	.037	.031
	(.187)	(.178)	(.187)	(.185)
Black	.023	-.147	.015	.128
	(.263)	(.256)	(.263)	(.257)
Hispanic	-.015	-.370	-.050	.388
	(.286)	(.289)	(.289)	(.270)
Asian	-.414	-.505	.658	.100
	(.407)	(.389)	(.334)**	(.358)
Married	.154	.053	-.136	-.053
	(.117)	(.115)	(.117)	(.116)
College Educ.	.022	-.098	.054	.027
	(.149)	(.145)	(.151)	(.149)
N	2227	2227	2227	2227
chi^2	6.297	4.788	6.928	6.157
p value	.614	.78	.544	.63
pseudo R^2	.003	.002	.003	.002

Table 3: Effect of Treatment Assignment on Free Trade Responses

International Agreement	-.086
	(.148)
Economic	.059
	(.144)
Placebo	.066
	(.146)
N	1417
chi^2	1.381
p value	.71
pseudo R^2	.0005

Table 4: Approval Rates by Treatment Group

Treatment Group	N	Proportion Approv.	Difference	SE	t stat	p value
Null	529	0.688				
Int. Agr.	519	0.580	-0.108	0.030	-3.65	<0.01
Econ	542	0.675	-0.013	0.028	-0.45	0.653
Placebo	538	0.645	-0.043	0.029	-1.49	0.136

Table 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	191	0.576				
Int. Agr.	189	0.524	-0.052	0.051	-1.02	0.309
Econ	198	0.586	0.010	0.050	0.20	0.843
Placebo	189	0.550	-0.026	0.051	-0.50	0.615

No Opinion Respondents

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	83	0.735				
Int. Agr.	79	0.595	-0.140	0.074	-1.90	0.059
Econ	92	0.674	-0.061	0.069	-0.88	0.381
Placebo	82	0.659	-0.076	0.072	-1.06	0.288

Protectionist Respondents

Treatment Group	N	Proportion Approv.	Difference	SD	t stat	p value
Null	62	0.952				
Int. Agr.	54	0.889	-0.063	0.050	-1.21	0.211
Econ	67	0.881	-0.071	0.049	-1.44	0.151
Placebo	67	0.866	-0.086	0.051	-1.68	0.095

Table 6: Approval Rates by Treatment Group: Primed vs. Unprimed Respondents

Treatment Group	N	Proportion Approv.	Difference	SE	t stat	p value
$\sim$ Primed: Null	121	0.702				
$\sim$ Primed: Int. Agr.	121	0.603	-0.099	0.061	-1.62	0.106
Primed: Null	114	0.658				
Primed: Int. Agr.	112	0.634	-0.024	0.064	-0.38	0.708

Table 7: Approval Rates by Treatment Group: Why Support International Agreements?

Treatment Group	N	Proportion Approv.	Difference	SE	t stat	p value
Null	115	0.687				
Consistency	118	0.576	-0.102	0.063	-1.61	0.109
Retaliation	121	0.554	-0.125	0.063	-1.97	0.050
Legality	116	0.560	-0.118	0.064	-1.85	0.065