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

Stephen Chaudoin

Abstract 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 actual policy choices. However, audiences also have strong preferences over the policy choices themselves, regardless of their consistency with past commitments. I conducted 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 constituents' underlying preferences limit the degree to which audience costs influence policymakers' calculations.

According to audience costs theory (ACT), audiences punish policymakers for committing to one policy and then renegeing on that promise. This theory has been frequently applied to international cooperation and crisis bargaining. In the context of crisis bargaining, promises are benchmarked against a leader's verbal statements. In the context of international cooperation, however, promises are benchmarked against the terms of legal agreements. Policymakers commit to certain policies when they ratify international agreements or join an institution. ACT argues that audiences punish policymakers who choose noncompliant policies that contravene their international obligations. These *ex post* audience costs facilitate coop-

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, Soo Yeon Kim, Jeffrey Kucik, John Londregan, Helen Milner, Daniel Nielsen, Emily Ritter, Tom Scherer, and Dustin Tingley, as well as the advice from the editors of *International Organization* and anonymous reviewers. I am especially grateful to the Niehaus Center for Global Governance that provided support for this research.

International Organization 68, Winter 2014, pp. 235–56

© 2014 by The IO Foundation.

doi:10.1017/S0020818313000386

eration and enhance the credibility of commitments by making compliance more attractive *ex ante*.¹

While the term *audience costs* encompasses many meanings, a key feature is that audiences have preferences over *consistency*. Audiences care whether a policymaker's actions are consistent with past promises. Fearon argues that inconsistency allows domestic political opponents to criticize the incumbent for damaging the country's international "credibility, face or honor."² Smith argues that audiences punish inconsistency because breaking commitments signals a leader's incompetence.³ In the context of international 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.⁴

However, audiences also have preferences over *policy* that have nothing to do with a course of action's consistency with past agreements. Consider workers who stand to lose their jobs if their elected representatives lower tariffs on certain imports. Even if those tariffs violate free trade agreements, the workers are unlikely to support a policy of lower tariffs. In other words, the workers' preferences over policy (high tariffs preferred to low tariffs) trump their preferences over consistency (high tariffs are inconsistent with prior commitments, while low tariffs are consistent).

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

Audience reactions to policymaker decisions over international cooperation have two components: a consistency effect—a negative reaction to policies that diverge from past agreements—and a policy effect—a reaction to divergence from the audience's preferred policy. Understanding the relative magnitude of the two effects is important for evaluating how international agreements and institutions affect member state behavior. If consistency effects are strong, as ACT theorizes, this is

1. See Leeds 1999; and Mansfield and Pevehouse 2006.

2. Fearon 1994, 581.

3. Smith 1998. Ashworth and Ramsay 2010 derive conditions under which audiences impose costs as part of an optimal incentive scheme. See also Snyder and Borghard 2011.

4. Abbott and Snidal 1998, 428.

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, institutions' effects 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 circumscribed.

To distinguish between consistency and policy effects, I embedded an experiment in a survey conducted in May 2012. The survey consisted of two parts. The first part presented respondents with a hypothetical situation regarding a policymaker's decision 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 and Levendusky and Horowitz, this captures the effects of consistency on approval of policymaker decisions.⁵

The second part of the survey asked respondents about their trade policy preferences. This allows me to examine whether, and to what degree, the respondent's policy preferences moderate consistency effects, showing whether treatments based on consistency have a stronger or weaker effect depending on the respondent's policy preferences. Like previous studies, I initially find strong consistency effects. When respondents are told that their leader's policies were inconsistent with international agreements, their approval of their leader's actions decreases significantly. However, unlike previous research, this effect is present for only respondents who do not already hold strong policy preferences. For respondents with strong preferences over the policy in question, informing them of that inconsistency has a significantly smaller effect.

These findings suggest that policy preferences are a stronger explanation of audience reactions than preferences over consistency. As a result, leaders choosing policy are more constrained by their audience's policy preferences than by their past commitments. Institutions and agreements, through their potential to activate audiences who prefer consistency, are likely to have weaker effects for countries with audiences 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. A key challenge facing international institutions is not simply to provide information or awareness about leaders who violate their international obligations, but also to persuade stubborn audiences who do not support compliance in the first place.

5. See Tomz 2007 and 2008; and Levendusky and Horowitz 2012.

Preferences Over Consistency and Policy

ACT argues that domestic populations punish leaders who make commitments to policies or courses of action and then choose policies that are inconsistent with those commitments.⁶ Audience costs have alternatively been described as “the surge in disapproval that would occur if a leader made commitments and did not follow through,”⁷ and “the punishments, in the form of lower support, meted out by domestic populations against leaders that make foreign threats but then ultimately back down.”⁸ 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.

This theory has been applied to both the context of crisis bargaining and international cooperation. In the context of international cooperation, ACT hypothesizes that leaders who break their international agreements suffer audience costs, which makes compliance with an agreement more attractive than defection. A leader’s policies are compared with the policies prescribed or proscribed by a particular international agreement. Noncompliant policies trigger backlash, which creates a strong disincentive for a leader contemplating defection from their international obligations.⁹

Audiences also have preferences over the deeds themselves, regardless of their consistency with past promises. Audience members assessing the leader’s performance might care about the leader’s consistent promises and policies, but they also have preferences over their leader’s actual actions. In virtually every issue concerning international cooperation, there are groups within countries who support and oppose the policies prescribed by an agreement. 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. The perceived or actual effects of trade policy adjustments have been linked to many features of trade policymaking.¹⁰ A rich body of literature examines variation in support for European integration across and within countries and variation in support for environmental cooperation.¹¹ In the highly charged context of human rights and war crimes, there is significant variation within countries over whether to legally try or punish current or recently removed leaders. Audience members’ support for the accused politician strongly tempers their preferences over whether that politician should be punished.

6. Fearon 1994.

7. Tomz 2007, 823.

8. Levendusky and Horowitz 2012, 324.

9. Simmons 2010.

10. See Rogowski 1987; Hiscox 2002; Milner and Tingley 2011; and Margalit 2011.

11. For example, see Gabel 1998; and Kelemen and Vogel 2010.

Existing Micro-Level Evidence of Consistency Effects

The two most well-known empirical studies of the micro-foundations of audience costs, from Tomz and from Levendusky and Horowitz,¹² were designed to detect consistency effects in the context of crisis bargaining. 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 learned that the US leader threatened military action against the aggressor if it invaded; the aggressor invaded; and the United States did not follow through with its threat. In other words, the treatment group learned that their leader's words and deeds were inconsistent. Participants assigned to the control group were told that the US 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.¹³ In Tomz and Levendusky and Horowitz, respondents in the treatment group were approximately 16 and 22 percent more likely to disapprove of their president, respectively.¹⁴

This approach, however, involves two differences between the treatment and control groups—one pertaining to consistency and one pertaining to an important policy decision. The first difference is the one the survey design desires. 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: hawks and doves. Hawks might prefer a policy of making threats and support subsequent military action. If told that the president threatened but took no action, hawks may disapprove because they preferred military action, regardless of their preferences over commitment-policy consistency. Doves might strongly dislike both threats to use force and military action. If told that the president threatened and backed down, doves may disapprove because of their dislike of threats. In other words, a hawk or a dove might have a very different counterfactual in mind than the one implied by the

12. See Tomz 2007; and Levendusky and Horowitz 2012.

13. 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, such as the level of escalation or the predicted amount of US casualties involved with following through on the threat, affect the magnitude of audience costs. Levendusky and Horowitz 2012 analyze whether domestic factors, such as the party of the president and congressional majorities match or the president's justification for backing down, affect the magnitude of audience costs.

14. *Ibid.*

survey design. Rather than comparing “threat plus backing down” (treatment scenario) with “no threat” (control scenario), they may be comparing the treatment scenario with their ideal scenario. For a hawk, this is “threaten plus invasion,” and for a dove, “no threat or invasion.” This difference between treatment and control groups creates the possibility that disapproval stems from the respondent’s dislike of inconsistency, of policy, or of both.

Two additional studies use approaches more closely resembling the one I used. Tomz analyzes a survey where respondents were randomly assigned to treatments that consisted of different arguments for or against an embargo on imports from Burma.¹⁵ Respondents who were given an argument against the embargo—that it violated international law—were 17 percent more likely to oppose the embargo than respondents who did not receive this argument. In related American politics research, Tomz and Van Houweling examine how voters respond to candidates who change their stance on an issue.¹⁶ Their survey research analyzes *valence* and *proximity* effects of candidate repositioning on voter opinions. 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, that is, a proximity effect. Tomz and Van Houweling use experiments where respondents read about candidates’ positions on taxes and abortion over time.¹⁷ They find strong evidence of valence effects, which are more moderate for voters who 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.

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 dislike of inconsistency and how much stems from preferences over the particular policy chosen? To answer this question, I embedded a randomized experiment in a survey conducted in May 2012. Respondents were randomly assigned treatment consisting of pros and cons of a particular policy, with one con arguing that the policy was inconsistent with past promises—namely an international agreement. Respondents were told that their politician enacted that policy and were asked whether they approved or disapproved of the politician’s actions. Respondents were then 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.

15. Tomz 2008.

16. Tomz and Van Houweling 2012.

17. Ibid.

The experiment's goal was 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 international agreements; 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, which has a larger effect on respondents' approval of a certain policy: their *ex ante* preferences regarding that policy or the experimental treatment in which they potentially learn that the policy is inconsistent with past agreements?

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 possess latent opposition to a policy, learning of its inconsistency with past promises should have minimal effect. Treatments pertaining to inconsistency simply provide "yet another" reason to oppose the policy and 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. If the respondent supports that policy for a variety of reasons unrelated to its consistency or inconsistency with past agreements, then learning about inconsistency has to overcome those initial contributors to the respondent's support to move her approval level significantly. Learning about inconsistency should have the strongest effect for respondents without strong preferences over the underlying policy.

The experiment was conducted in the context of trade policy and described the president's decision over whether to impose tariff barriers against certain imports. Extant literature describes many reasons why respondents might have *ex ante* preferences over import barriers that have nothing to do with potential inconsistency between tariff barriers and past trade agreements. Respondents who expect to lose from increased trade because of their factor endowments, their sector of employment, or their negative perceptions of the overall effect of free trade on the economy already have strong reasons to oppose tariff barriers. Telling such respondents that tariff barriers are inconsistent with past agreements may make them less supportive of that policy, but this effect is likely to be marginal since the respondent's dislike of inconsistency has to pull against these other interests. Respondents who oppose tariff barriers for the opposite reasons should also display smaller effects of being told that tariffs are inconsistent with past agreements. Approval of a politician enacting such a policy can only drop so low. If approval is already low, it is harder for the inconsistency treatment to push it much further downward.

Survey Recruitment

Approximately 2,500 survey respondents were recruited using Amazon's Mechanical Turk (mTurk) service. mTurk is an online web-based platform where researchers can post "tasks" and compensation levels for participants who complete the

tasks. In this case, the task was to complete the survey. Compensation ranged from \$0.55 to \$0.75 per respondent for a survey taking approximately ten minutes. Participants who accepted the task on mTurk were directed to an external survey site (Qualtrics), answered survey questions, and were given a unique code to enter to receive their compensation from Amazon.¹⁸ Berinsky, Huber, and Lenz show that subjects recruited on mTurk are more representative of the US population than convenience samples, though less representative than subjects recruited via nationally representative Internet-based samples or national probability samples.¹⁹ The respondent pool was relatively close to nationally representative surveys, though, unsurprisingly for an Internet-based survey, respondents were 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 percent male, compared to 42.8 percent male in the 2008 ANES), more liberal,²⁰ and more likely to never have been married (35.2 percent compared to 14.2 percent in the 2008 ANES).

Main Experiment

For the main experiment, respondents were presented with a hypothetical situation involving a fictional US company, Arena, Inc. This company manufactured metal brackets, which, as respondents were told, US construction companies used in construction. Respondents were then told that a European company had recently begun producing similar brackets at a lower price, and that US construction companies had begun buying the foreign brackets instead of the US-produced brackets. I left the foreign 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 perceptions of more politically charged import partners such as 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 US 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 listed below—arguments opposing import restrictions—or a null treatment in which the respon-

18. I owe appreciation to Peer et al. 2012, who provide a useful script for ensuring that respondents do not take the survey more than once. For a fuller description of the mTurk platform, see Berinsky, Huber, and Lenz 2012.

19. Berinsky, Huber, and Lenz 2012.

20. The mean political ideology score on a seven-point scale (7 = strong conservative) for this survey was 3.41, compared with 4.30 in the 2008 ANES.

dent was not given a con.²¹ To avoid stacking the deck in favor of finding effects for any one treatment, each has identical sentence structures as well as similar word counts and tone.

- 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 United States and Europe, and Europe would sue the United States 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 US construction companies have to buy more expensive US 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 is that import restrictions are contrary to a previous commitment, an inconsistency that could 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—when a country violates its agreement, a supra-national judicial body can be called upon to condemn those defections.²²

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 import restrictions might save some jobs, but would also likely cost other jobs. The placebo treatment matches the other two treatments in word count and structure, but does not contain any specific content.

After receiving the standard pro-import restriction argument and one of the four treatments (IA, economic, placebo, null), 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 president handled the situation, and could answer: “Strongly approve,” “Somewhat approve,” “Neither

21. Hiscox 2006 shows in a similar context that antitrade priming significantly affects survey respondents’ attitudes toward increasing trade.

22. To be sure, respondents could react to the IA treatment for various reasons—dislike of inconsistency, loss of national honor or reputation, or updated beliefs about the leader’s quality. I conducted follow-up surveys analyzing why respondents care about violations of international agreements. Results and analysis are available in the online appendix at the author’s Dataverse site at (<http://dvn.iq.harvard.edu/dvn/dv/chaudojn/>).

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 and Levendusky and Horowitz.²³ I constructed a binary variable measuring approval versus disapproval, 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 approval 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. The null treatment provides a 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 assess how much of any treatment effect comes from the specific content of the treatment, and how much comes from the fact that the respondent was given simple words on the page that were opposed to the policy (placebo treatment). It is possible that respondents simply count pros and cons, so having any arguments listed as a con increases disapproval, regardless of content. Comparing the effects of the placebo treatment with the IA and economic treatments helps isolate the additional effect on approval caused by the specific content of those treatments.

Other Questions and Survey Balance

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 standard demographic questions. Embedded in the post-experiment series was a question pertaining directly to trade policy.

I first checked that treatment assignment was not skewed among the covariates measured from these pre- and postexperiment questions. For each of the four treatment groups, there is no evidence of imbalance in the pretreatment covariates, using the test from Hansen and Bowers.²⁴ The overall χ^2 statistics and associated *p*-values for each treatment group areas follows: IA, 4.86 (*p* = 0.772); economic, 5.81 (*p* = 0.669); placebo, 6.30 (*p* = 0.614); and null, 6.05 (*p* = 0.642). Even including all pre- and posttreatment covariates, there is no strong evidence of imbalance.²⁵

23. See Tomz 2007; and Levendusky and Horowitz 2012.

24. Hansen and Bowers 2008.

25. For a full description, see online appendix. The overall χ^2 statistics and associated *p*-values for each treatment group, using all covariates, are as follows: IA, 17.3 (*p* = 0.505); economic, 25.3 (*p* = 0.118); placebo, 13.6 (*p* = 0.752); and null, 13.8 (0.741). For the economic treatment, approximately 61 percent of the treatment group was employed, compared to 54 percent in the control group. This potentially biased the economic treatment effect downward but was unlikely to have had strong

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 they had received in the main experiment from a list of four possible arguments. Almost 86 percent (85.9) correctly recalled that they had been given a pro-import restriction argument pertaining to layoffs by the US firm, among a list containing the correct answer and two fabricated arguments in favor of import restrictions. Sixty-two percent (62.1) correctly recalled the anti-import restriction they had been given (if any) from a list containing the four possible treatments. The placebo treatment, unsurprisingly, was the weakest, with only 49.3 percent of respondents correctly recalling it. The IA and economic treatments were stronger, with 66.7 and 68.5 percent, respectively, correctly recalling the con arguments they had been given. Of respondents who received the null treatment, 63.8 percent 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, that is, binomial tests that the proportion of correct responses equaled 0.33 (for the pros) or 0.25 (for the cons), at the 0.01 level.

Trade Policy Preferences

To measure policy preferences, I also asked a standard free trade question in the middle of the postexperiment questions. 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, 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.²⁷

Eliciting respondents’ preferences over free trade lets me compare the relative magnitudes of consistency effects and policy preference effects. I can also compare the magnitude of treatment effects across respondents with different policy preferences. To the degree that policy preferences moderate consistency effects,

effects on the analysis, which focuses on the IA treatment. Tests conducted using RIttools in R; Bowers, Fredrickson, and Hansen 2010.

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

27. See online appendix for follow-up surveys where preferences are primed rather than elicited.

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 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 treatment assignment "contaminated" respondents' answers to the free trade question. The survey was designed to dampen such effects by placing a large number of questions between the main experiment and trade policy question. There is no strong evidence that the treatment each respondent received affected responses 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. None of the treatment assignments had a significant effect on the probability of a respondent being pro-free trade, protectionist, or having no opinion.²⁸

Experimental Results

Figure 1 and Table 1 show the percentage of respondents who approved of presidents who implemented import restrictions across each of the treatment groups.²⁹ Among those who received the null treatment, 68.8 percent approved of the

28. Full results are in the online appendix. 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. These (null) results are robust to alternate specifications, such as multinomial regressions or difference in means tests for trade policy responses across treatment groups.

29. 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 {null, IA, economic, 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 noninformative Jeffrey's prior. 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 percent credibility intervals are from 5,000 draws from the posterior distributions.

president's actions. Among those receiving the IA treatment, only 58 percent approved of the president's actions. The difference between the null and IA approval rates measures the strong drop in approval that occurs when respondents learn that their president's actions violated prior agreements. Approval rates are 10.8 percent lower in the IA treatment group than in the null group. This difference is highly statistically significant (p -value for the difference in proportion approving is <0.01).³⁰

TABLE 1. Approval rates by treatment group

Treatment group	N	Proportion approval	Difference	Standard deviation	t-stat	p-value
Null	529	0.688				
International agreement	519	0.580	-0.108	0.030	-3.65	<0.01
Economic	542	0.675	-0.013	0.028	-0.45	0.653
Placebo	538	0.645	-0.043	0.029	-1.49	0.136

The other two treatments do not have significant effects on approval rates. The economic treatment decreased approval slightly relative to the null group, to 67.5 percent. Even direct economic concerns, such as the possibility of job loss in other industries, do not appear to influence approval rates. For the placebo treatment, 64.5 percent approved, a 4.1 percent 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 policy and past agreements 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 expressed preferences regarding 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 2.

For pro-free-trade respondents and protectionist respondents, the difference between approval rates for the null group and the IA group are small and insignificant. Among pro-free-trade respondents, approval rates in the IA treatment group were 52.4 percent compared to 57.6 percent for the null treatment group. The dif-

30. The standard deviation, t -statistic, 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.

ference (-5.2 percent) is approximately half the difference found for the entire sample (-10.8 percent), and is statistically insignificant (p -value = 0.309). Among protectionist respondents, approval rates for the IA treatment group were 88.9 percent 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).

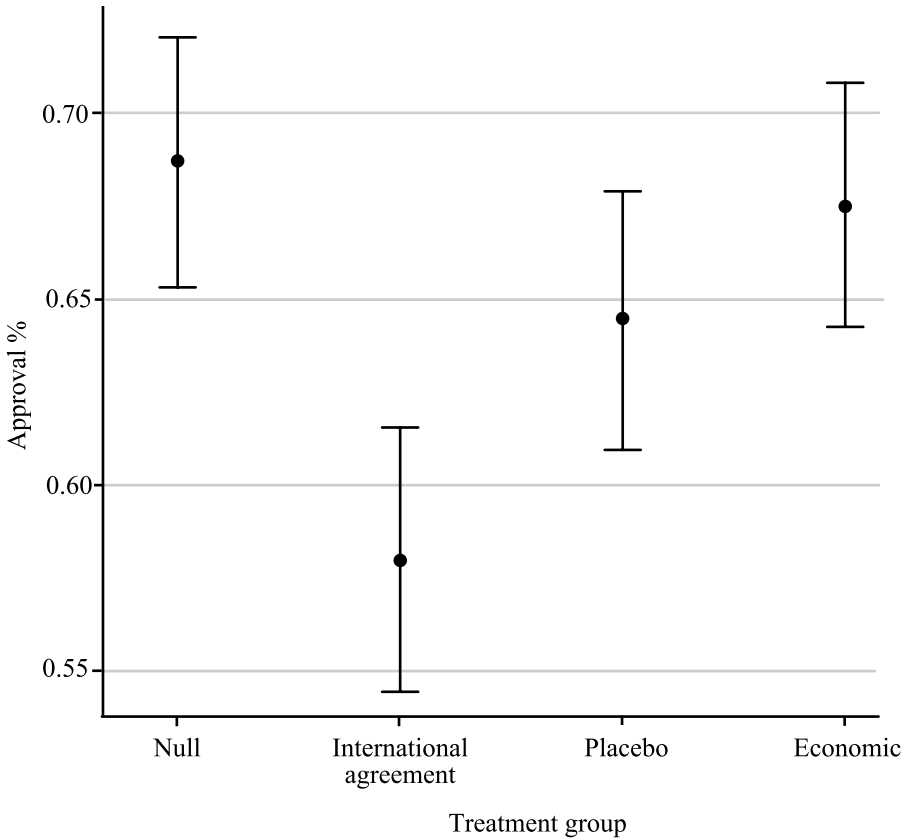


FIGURE 1. *Treatment effects, all respondents*

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, approval rates in the IA group were 59.5 percent, compared to 73.5 percent for the null group. The difference of -14.0 percent is substantively large and statistically significant (p -value = 0.059). Consistency effects

are strongest for respondents without strong policy preferences, and 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 had a significant effect on only those respondents who did not hold strong opinions over free trade in general. Stated 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 the underlying aversion to free trade. Respondents without strong opinions on free trade were the most malleable, and most influenced by inconsistency between policy and agreements.

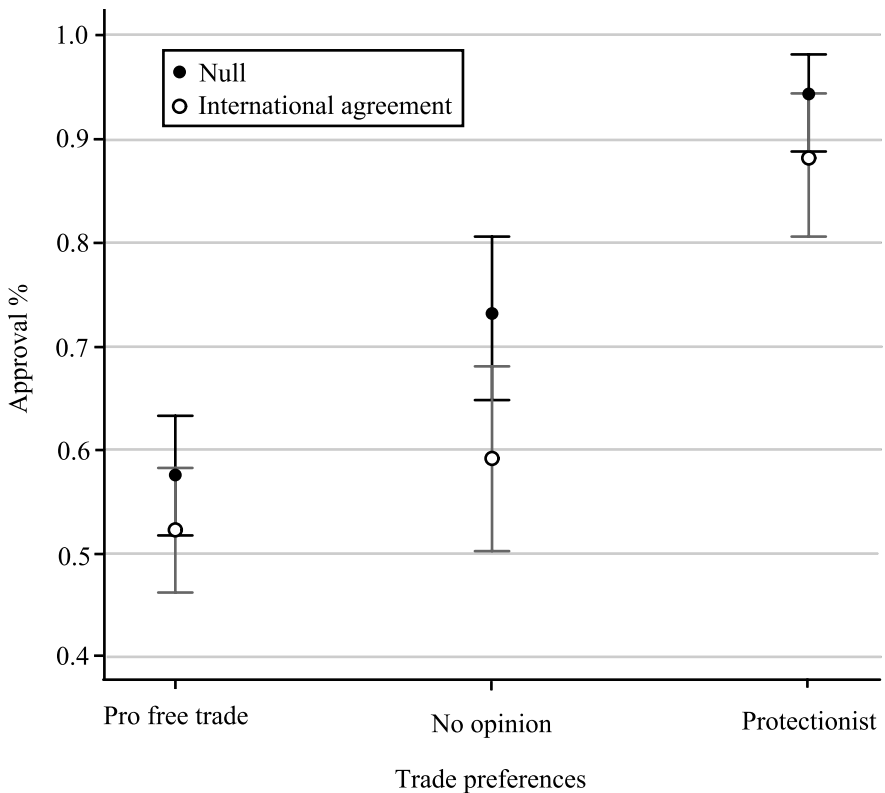


FIGURE 2. *International agreement treatment effects, by respondent trade preferences*

These results are consistent with Tomz and Van Houweling's analysis of domestic tax and abortion policy.³¹ They find that valence (consistency) effects are strongest among respondents who do not consider the issue important. Among respondents who considered tax or abortion policy particularly important, proximity (policy) effects were most important. If the respondent cared strongly about the issue, then 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.

TABLE 2. *Approval rates by treatment group and by respondent trade preference*

<i>Pro-free trade respondents</i>						
<i>Treatment group</i>	<i>N</i>	<i>Proportion approval</i>	<i>Difference</i>	<i>Standard deviation</i>	<i>t-stat</i>	<i>p-value</i>
Null	191	0.576				
International agreement	189	0.524	-0.052	0.051	-1.02	0.309
Economic	198	0.586	0.010	0.050	0.20	0.843
Placebo	189	0.550	-0.026	0.051	-0.50	0.615
<i>No-opinion respondents</i>						
<i>Treatment group</i>	<i>N</i>	<i>Proportion approval</i>	<i>Difference</i>	<i>Standard deviation</i>	<i>t-stat</i>	<i>p-value</i>
Null	83	0.735				
International agreement	79	0.595	-0.140	0.074	-1.90	0.059
Economic	92	0.674	-0.061	0.069	-0.88	0.381
Placebo	82	0.659	-0.076	0.072	-1.06	0.288
<i>Protectionist respondents</i>						
<i>Treatment group</i>	<i>N</i>	<i>Proportion approval</i>	<i>Difference</i>	<i>Standard deviation</i>	<i>t-stat</i>	<i>p-value</i>
Null	62	0.952				
International agreement	54	0.889	-0.063	0.050	-1.21	0.211
Economic	67	0.881	-0.071	0.049	-1.44	0.151
Placebo	67	0.866	-0.086	0.051	-1.68	0.095

This pattern is also displayed when considering the economic and placebo treatments shown in Figure 3. Respondents with established opinions on free trade

31. Tomz and Van Houweling 2012.

were less moved by either treatment. Respondents without strong opinions on free trade were more influenced by both treatments. The economic treatment actually has a positive (though small and insignificant) effect on approval rates among pro-free-trade respondents, 1 percent. It has a larger and negative effect among no-opinion and protectionist respondents, -6.1 percent and -7.1 percent respectively, though both are insignificant.

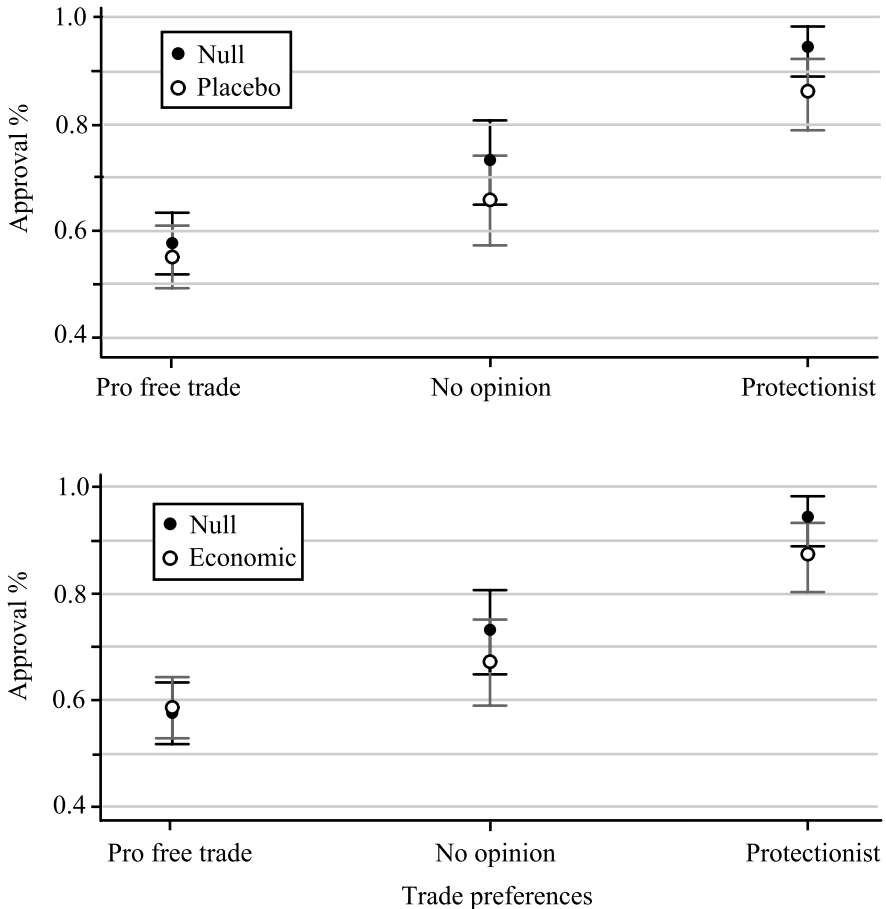


FIGURE 3. *Placebo and economic treatment effects, by respondent trade preferences*

Among pro-free-trade respondents, the difference between approval rates in the placebo and null groups was insignificant. Yet for respondents expressing no opin-

ion, the placebo treatment managed to decrease approval by 7.6 percent, though this difference falls short of conventional significance levels (p -value = 0.288).³²

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. Limiting analysis to only the respondents where I found a significant IA treatment effect, the IA treatment effect was statistically indistinguishable from the placebo treatment effect. Both the IA and placebo treatments induce lowered approval relative to the null treatment. But for no-opinion respondents, approval rates are only 6.3 percent lower under the IA treatment than under the placebo treatment, and this difference is statistically insignificant (p -value = 0.407). While I can confidently say that both the IA and placebo treatments have distinct effects, I can less confidently say that the IA treatment has a distinct effect 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, the respondent's preferences over the policy itself is a better predictor than whether the respondent knows the policy is inconsistent with the leader's previous commitments. Using ordinary least squares (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.

Conclusion

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 or agreements. This study examined how much a voter's approval of a leader's policy is driven by preferences over the consistency between that policy and past commitments and how much it is driven by preferences over the policy itself.

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 in their approval of leaders. How-

32. Protectionist respondents were also influenced by the placebo treatment, which decreased approval relative to the null by 8.6 percent, with an associated p -value of .095.

ever, 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. Those costs are significantly moderated for groups with policy opinions.

The finding that audience costs are moderated by preferences over policy is bad news for ACT as applied to the question of how international agreements facilitate cooperation. To the degree that audience preferences over consistency trump preferences over policy, then ACT predicts a robust, consistent effect of international commitments on member state behavior. Agreements are strong forces for compliance because, for leaders choosing whether to comply with that agreement, 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, leaders' commitment acts as a strong inducement to honor their promise by cooperating.

However, to the degree that preferences over policy endure, even after leaders have made commitments, the effects of audience costs are weaker. 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 previous commitments is similar to the calculus facing a leader who has made commitments. In both "worlds," the leader's decision calculus is largely based on the expressed or anticipated audience preferences over policy. 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. The marginal effect of audience costs on leaders' calculations decreases.

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 effect on changing public opinion and in turn, less effect on policymakers beholden to those groups.

This question is likely to be particularly 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 these areas, 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 that, though important to subsets of the population, are not as salient 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 to these rules. Others, such as consumers who potentially benefit from compliance via lower prices, 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 to use military force to back up a threat. The use of 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, and by implication, whether their leader's commitments have been honored. This is in contrast with the context of international cooperation where many issue areas are governed by opaque policies, and compliance is difficult for audiences to observe. For example, audiences lack information about whether their government's emission-reductions efforts will meet international targets. In international trade, nontariff barriers are especially inaccessible for the average audience member, with democracies often deliberately obscuring their policies.³³ 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 survey results also suggest 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 containing no argumentative content were persuasive. This further dampens audience costs, 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 those arguments' validity. Levendusky and Horowitz find that audience costs are significantly lessened when the president claims that his actions were justified by new informa-

33. Kono 2006.

tion.³⁴ It is highly unlikely that a policymaker would ever break a prior promise 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 “How to persuade the intransigent?” A task for future research is to determine how international institutions and agreements can persuade domestic audiences who have a strong stake in noncompliance 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. *Journal of Conflict Resolution* 42 (1):3–32.
- Ashworth, Scott, and Kristopher W. Ramsay. 2010. Should Audiences Cost? Optimal Domestic Constraints in International Crises. Working paper. Chicago/Princeton, NJ: University of Chicago/Princeton University.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk. *Political Analysis* 20 (3):351–68.
- Bowers, Jake, Mark Fredrickson, and Ben Hansen. 2010. RIttools: Randomization Inference Tools. R package version 0.1-11.
- Fearon, James D. 1994. Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88 (3):577–92.
- Gabel, Matthew. 1998. Public Support for European Integration: An Empirical Test of Five Theories. *Journal of Politics* 60 (2):333–54.
- Hansen, Ben B., and Jake Bowers. 2008. Covariate Balance in Simple, Stratified and Clustered Comparative Studies. *Statistical Science* 23 (2):219–36.
- Hiscox, Michael J. 2002. *International Trade and Political Conflict: Commerce, Coalitions, and Mobility*. Princeton, NJ: Princeton University Press.
- . 2006. Through a Glass and Darkly: Attitudes Toward International Trade and the Curious Effects of Issue Framing. *International Organization* 60 (3):755–80.
- 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):427–56.
- Kono, Daniel Y. 2006. Optimal Obfuscation: Democracy and Trade Policy Transparency. *American Political Science Review* 100 (3):369–84.
- Leeds, Brett Ashley. 1999. Domestic Political Institutions, Credible Commitments, and International Cooperation. *American Journal of Political Science* 43 (4):979–1002.
- Levendusky, Matthew S., and Michael C. Horowitz. 2012. When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs. *Journal of Politics* 74 (2):323–38.
- Mansfield, Edward D., and Diana C. Mutz. 2009. Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety. *International Organization* 63 (3):425–57.

34. Levendusky and Horowitz 2012.

- Mansfield, Edward D., and Jon C. Pevehouse. 2006. Democratization and International Organizations. *International Organization* 60 (1):137–67.
- Margalit, Yotam. 2011. Costly Jobs: Trade-Related Layoffs, Government Compensation, and Voting in US Elections. *American Political Science Review* 105 (1):166–88.
- 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 (1):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, Carnegie Mellon University, Pittsburgh, PA.
- Rogowski, Ronald. 1987. Political Cleavages and Changing Exposure to Trade. *American Political Science Review* 81 (4):1121–37.
- Simmons, Beth A. 2010. Treaty Compliance and Violation. *Annual Review of Political Science* 13:273–96.
- Smith, Alastair. 1998. International Crises and Domestic Politics. *American Political Science Review* 92 (3):623–38.
- Snyder, Jack, and Erica D. Borghard. 2011. The Cost of Empty Threats: A Penny, Not a Pound. *American Political Science Review* 105 (3):437–56.
- Tomz, Michael. 2007. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization* 61 (4):821–40.
- . 2008. Reputation and the Effect of International Law on Preferences and Beliefs. Working paper. Stanford, CA: Stanford University.
- Tomz, Michael, and Robert P. Van Houweling. 2012. Candidate Repositioning. Unpublished manuscript, Stanford University, Stanford, CA.