To: Oliver E. Williamson Seminar on Institutional Analysis
From: Michael Tomz, Stanford University (tomz@stanford.edu)
Re: The Credibility of International Commitments
Date: January 2007

On January 25, I will present some of my research about the credibility of international commitments. This memo introduces the two short papers I have attached, and puts them in the context of my broader research agenda.

Every day, leaders make commitments to foreign governments, firms, and individual investors. In the absence of third-party enforcement, why does anyone trust or honor these cross-border commitments? My forthcoming book moves us closer to an answer, by showing that reputation has been key to understanding cooperation between sovereign governments and foreign investors over the past three centuries.[1]

The work I would like to discuss with you on the 25th takes a further step, by examining strategies to making reneging on international commitments more costly. One strategy involves *publicity*: publicizing international threats or promises before a domestic audience that could hold leaders accountable for hurting the country's reputation. A second strategy involves *legalization*: embedding commitments in international treaties that, though unenforceable, may be costly to break for reputational or normative reasons.

I am using both survey-based experiments and non-experimental methods to study these hand-tying strategies. The attached papers illustrate the experimental strain in my research. The paper entitled "Domestic Audience Costs in International Relations" is now forthcoming in *International Organization*, but I should be able to incorporate your comments into the final version. The paper entitled "The Effects of International Agreements on Foreign Policy Preferences" is more preliminary. I will discuss both papers during the seminar. Time permitting, I will also summarize experiments my students and I recently administered to members of the British Parliament and describe software we are developing to facilitate experiments in research and teaching.

This work is part of a larger project, which is supported by a five-year CAREER grant from the National Science Foundation. I am still in the first year of the grant, precisely the moment when advice from the Haas community could make a major difference. I look forward to your comments and suggestions!

---

[1] Michael Tomz, *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton University Press, forthcoming 2007.

Domestic Audience Costs in International Relations: An Experimental Approach

Michael Tomz
Stanford University
tomz@stanford.edu

October 2006

Forthcoming, *International Organization* 61, no. 4 (Fall 2007).

Abstract: What makes international threats credible? Recent theories point to domestic audience costs: the domestic price a leader would pay for making foreign threats and then backing down. This article provides the first direct evidence of audience costs. The analysis, based on experiments embedded in public opinion surveys, shows that audience costs exist across a wide range of conditions and increase with the level of escalation. The costs are evident throughout the population, and especially among politically active citizens who have the greatest potential to shape government policy. Finally, preliminary evidence suggests that audience costs arise because citizens care about the international reputation of the country or leader. These findings help identify how, and under what conditions, domestic audiences make commitments credible. At the same time, they demonstrate the promise of using experiments to answer previously intractable questions in the field of international relations.

Acknowledgements: I thank Time Sharing Experiments in the Social Sciences and the National Science Foundation (CAREER Grant SES-0548285) for financial support. Colleagues at Knowledge Networks provided invaluable assistance in fielding the surveys. For helpful comments I am grateful to Jim Fearon, Page Fortna, Jon Krosnick, Skip Lupia, Diana Mutz, Ken Scheve, Ken Schultz, Alastair Smith, Paul Sniderman, and Jessica Weeks.

Every day world leaders make threats and promises. Without a world government that compels leaders to keep commitments, many of which would be costly to carry out, when and why should people take foreign leaders at their word? The answer may lie at the intersection of foreign affairs and domestic politics.

Recent models of international relations assume that leaders will suffer "domestic audience costs" if they make threats or promises and fail to follow through. Citizens, it is claimed, will think less of leaders who back down than of leaders who never commit in the first place. In a world with audience costs, the prospect of losing domestic support—or even office—would give leaders incentives to avoid making empty threats and promises. The concept of domestic audience costs is now central to theories about military crises, and researchers have incorporated similar ideas into models of alliances, economic sanctions, foreign trade, foreign direct investment, monetary commitments, interstate bargaining, and international cooperation more generally.[1]

Despite the prominence of audience costs in international relations theories, though, it remains unclear whether and when such costs exist in practice. Most empirical work on the topic is indirect. Fearon conjectured that audience costs are higher in democracies than in autocracies, and then explained why this would lead the two regime types to behave differently in world

---

[1] The seminal paper is Fearon 1994. See also Schultz 2001a and Smith 1998 on military crises; Gaubatz 1996 and Smith 1996 on alliances; Dorussen and Mo 2001 and Martin 1993 on economic sanctions; Mansfield, Milner, and Rosendorff 2002 on trade; Jensen 2003 on foreign direct investment; Broz 2002 on monetary commitments; Leventoglu and Tarar 2005 on interstate bargaining; and Leeds 1999, Lipson 2003, and McGillivray and Smith 2000 on the role of audience costs in international cooperation in general.

affairs.[2] Researchers have, therefore, checked for correlations between democracy and foreign policy.[3] But such tests have ambiguous implications; they do not reveal whether the effects of democracy stem from audience costs or from other differences between political regimes.

It would be ideal to study audience costs directly, perhaps by examining what happened to leaders who issued threats and then backed down. The problem, which international relations scholars widely recognize, is strategic selection bias.[4] If leaders take the prospect of audience costs into account when making foreign policy decisions, then in precisely the situations when citizens would react harshly against backing down, leaders will tend to avoid that path, leaving little opportunity to observe the public backlash. Audience costs will be largely invisible. It would seem, therefore, that a direct and unbiased measure of audience costs is beyond reach.

This article aims to solve the empirical conundrum. The analysis is based on a series of experiments embedded in public opinion surveys. In each experiment, the interviewer describes a military crisis. Some participants are randomly assigned to a control group and told that the president does not get involved. Others are randomly placed in a treatment condition in which the president escalates the crisis but ultimately backs down. All participants are then asked whether they approve of the way the president handled the situation. By comparing approval ratings in the “stay out” and “back down” conditions, we can measure the domestic reaction directly without strategic selection bias.

---

[2] Fearon 1994.

[3] See, e.g., Eyerman and Hart 1996; Gelpi and Griesdorf 2001; Partell and Palmer 1999; Prins 2003.

[4] Baum 2004; Schultz 2001b.

In the remainder of this article, I show that constituents do indeed disapprove of leaders who make international threats and then renege. I then explain why many leaders regard disapproval as a political liability. Finally, as a step toward deepening our theoretical as well as empirical understanding of audience costs, I investigate why citizens react negatively to empty threats.

1. Do Audience Costs Exist?

Two questions are fundamental to the literature on audience costs. Would constituents disapprove if their leader made false commitments, and by what means would disapproving citizens hold their leaders accountable? The first question is analytically prior and the focus of this article; complementary work by others examines the secondary question of accountability.[5] Throughout this article, I use the term audience costs as shorthand for the surge in disapproval that would arise if the leader made commitments and did not follow through.[6]

There is much speculation about whether audience costs exist at all. Some analysts hypothesize that empty commitments would provoke a negative public reaction.[7] Citizens, it is argued, reason that hollow threats and promises undermine the country's reputation in world affairs; that empty commitments are dishonorable and embarrassing; or that inconsistency is evidence of incompetence.

---

[5] For a recent review and important extension of the literature on accountability and audience costs in democracies and autocracies, see Weeks 2006.

[6] I adopt this terminology with the understanding that changes in approval will be more consequential in some political systems than in others.

[7] e.g. Fearon 1994; Guisinger and Smith 2002; Smith 1998.

Other analysts believe, however, that citizens would not disapprove of committing and backing down. They point out that some citizens pay little attention to foreign policy, and others focus on final outcomes, rather than the sequence of threats and promises *in medias res*.[8] Finally, even citizens who pay careful attention to every jab and parry may forgive leaders for making false commitments. After all, anyone who has played poker understands that bluffing can be an optimal strategy. "Why, then, would constituents punish leaders whose bluffs are sometimes called?"[9]

Citizens may even prefer leaders who try before conceding over leaders who forfeit at the outset. Walt, for example, points out that citizens may "reward a leader who overreaches at first and then manages to retreat short of war. Thus the British and French governments did not suffer domestic audience costs when they backed down during the Rhineland crisis of 1936 or the Munich crisis of 1938, because public opinion did not support going to war."[10] Walt's historical examples raise an interesting possibility: maybe leaders can gain points by escalating before giving up, instead of giving way immediately.

Do citizens typically respond with scorn, indifference, or praise when their leaders commit without following through? Until we know, we cannot understand the effects of publicly committing before a domestic audience. If audience costs do exist under general conditions, this discovery would provide—for the first time—empirical microfoundations for a broad class of models in international security and political economy. The discovery would also suggest

[8] Brody 1994, 210.

[9] Gowa 1999, 26. See also Desch 2002, 29-32; Ramsay 2004; Schultz 1999, 237; Slantchev 2006.

[10] Walt 1999, 34.

profitable avenues for new research, especially if the domestic reaction to flip-flopping varied systematically with characteristics of the situation and the audience. If, on the other hand, citizens showed no stronger preference for leaders who avoided commitments than for leaders who committed and subsequently reneged, we would need to rethink how leaders send signals and tie hands in world affairs.

2. Methods

To study audience costs directly while avoiding the problem of selection bias, I designed and carried out a series of experiments. The first experiment was administered to a nationally representative random sample of 1,127 U.S. adults in 2004. (Sampling methods are discussed in the Appendix.) All participants in the internet-based survey received an introductory script: "You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach U.S. leaders have taken, and ask whether you approve or disapprove."[11]

Participants then read about a foreign military crisis in which "A country sent its military to take over a neighboring country." To reduce the risk that idiosyncratic features of particular crises drove the results, I randomly varied four contextual variables—regime, motive, power, and interests—that have been shown to be consequential in the IR literature.[12] The country was led by a "dictator" in half the interviews and a "democratically elected government" in the other half. The attacker sometimes had aggressive motives—it invaded "to get more power and

[11] The full text of all experiments is available at www.webpage.edu.

[12] The literature on these four variables is vast. Herrmann and Shannon 2001 and Herrmann, Tetlock, and Visser 1999 discuss their impact on elite and mass opinion.

resources"—and sometimes invaded "because of a longstanding historical feud." To vary power, I informed half the participants that the attacker had a "strong military," such that "it would have taken a major effort for the United States to help push them out," and told the others that the attacker had a "weak military," which the United States could repel without major effort. Finally, a victory by the attacking country would either "hurt" or "not affect" the safety and economy of the United States.

Having read the background information, participants learned how the U.S. president handled the situation. Half the respondents were told: "The U.S. president said the United States would stay out of the conflict. The attacking country continued to invade. In the end, the U.S. president did not send troops, and the attacking country took over its neighbor." Remaining respondents received a scenario in which the president made a threat but did not carry it out. "The U.S. president said that if the attack continued, the U.S. military would push out the invaders. The attacking country continued to invade. In the end, the U.S. president did not send troops, and the attacking country took over its neighbor." The language in the experiment was purposefully neutral: it objectively reported the president's actions, rather than using interpretive phrases like "backed down" or "wimped out" or "contradicted himself," which might have biased the research in favor of finding audience costs.[13]

After displaying bullet points that recapitulated the scenario, I asked: "Do you approve, disapprove, or neither approve nor disapprove of the way the U.S. president handled the situation?" Respondents who approved or disapproved were asked whether they held their view very strongly, or only somewhat strongly. Those who answered "neither" where prompted: "Do

[13] The experiment also avoided language that might have reduced audience costs, either by criticizing the president who stayed out or by praising the leader who escalated the crisis.

you lean toward approving of the way the U.S. president handled the situation, lean toward disapproving, or don't you lean either way?" The answers to these questions implied seven levels of presidential approval, ranging from very strong disapproval to very strong approval.

By design, the experimental groups differed in only one respect: whether the U.S. president escalated the crisis before letting the attacker take over its neighbor. For this reason, any systematic difference in presidential approval was *entirely* due to the path the president took, not to variation in background conditions or the outcome of the crisis.

This experimental approach offers distinct advantages, but it is not infallible. Indeed, experiments are vulnerable on the dimension where observational data is most compelling: external validity. Do citizens behave differently in interviews than in actual foreign policy crises? If so, do the experiments in this article understate or overstate the magnitude of audience costs? It is hard to say for sure. Ultimately, the best way to make progress on complicated topics is to analyze data from multiple sources. As the first of their kind, the experiments in this paper provide new insights to complement what others have found with historical data.

3. Findings: Direct Evidence of Audience Costs

The experiments described above offered a new way to test competing conjectures in the literature. If audience costs exist, respondents who received the vignette in which the president stayed out should have approved more than respondents who read that the president threatened and yielded. If, on the other hand, citizens do not disparage leaders for getting caught in a bluff, levels of approval should be approximately the same in the two experimental groups. Finally, if leaders score points at home by showing at least some effort abroad, popularity should be higher in the "empty threat" scenario than in the "stay out" scenario.

Which of these conjectures best fits the data? Before answering that question, I confirmed that the treatment and control groups were balanced on baseline covariates that could affect presidential approval. Specifically, I used logistic regression to estimate whether any demographic or contextual variables predicted membership in the treatment group. Not one of the many variables in the model—gender, age, education, income, urban residence, political party identification, a history of military service, attitudes toward internationalism and the use of force, stakes for the United States, and the motive, power, and interests of the invader—had a statistically significant effect on the probability of receiving the "empty threat" vignette. Based on a likelihood ratio test, we cannot reject the hypothesis that the relationship between the treatment and *all* baseline variables was zero.[14]

Having established that the treatment was random, I examined how the public responded to each path the president traveled.[15] The results, in Table 1, provide unambiguous evidence of audience costs . For each presidential strategy, the table gives the percentage of respondents who disapproved, approved, or expressed an intermediate view. As the table shows, the president who issued an empty threat (column 1) was significantly less popular than the president who never got involved (column 2). Empty threats caused both strong and moderate disapproval to grow, and they led both moderate and strong approval to shrink. Evidently, backtracking on a

[14] The likelihood ratio test statistic, 10.18, was distributed chi-squared with 16 degrees of freedom. If all coefficients were zero, we would observe a test statistic that large with probability .86.

[15] Due to randomization, there is little need for elaborate statistical models with batteries of control variables. One can obtain unbiased estimates of the treatment effect via cross-tabulations.

verbal threat evoked an adverse public reaction. The political consequences of swings in approval—their effect on presidential power and incentives—are discussed later in the paper.

[TABLE 1 ABOUT HERE]

The final two columns of Table 1 summarize the magnitude of the effects. Compared to a baseline condition in which the president stayed out, the decision to threaten and not follow through caused disapproval to swell by 16 points, with a 95 percent confidence interval ranging from 10 to 22.[16] At the same time, the percentage of fence-sitters (citizens who neither approved nor disapproved) fell by 4 points, and the share of approvers dropped by 12 points, with an associated confidence interval from 8 to 17. Thus, a mere threat—without any military deployment or use of force—exposed the president to potentially widespread disapproval.

4. Do Audience Costs Increase with the Level of Escalation?

The previous experiment established that even mild acts of escalation—making verbal threats—can set the stage for substantial audience costs. I now investigate whether the public response increases with the level of hostility. If so, leaders can send progressively stronger signals by ratcheting crises to higher levels. The literature on militarized interstate disputes (MIDs) distinguishes three levels of escalation prior to war.[17] *Threats* are verbal indications of hostile intent, such as the ultimatum in the first experiment. The next rung on the escalatory

[16] I obtained confidence intervals by modeling each proportion as a beta distribution with a uniform prior, and then using random draws from each beta distribution to simulate the posterior of each proportion (and differences between proportions, and ratios of proportions). See Johnson and Albert 1999.

[17] Jones, Bremer and Singer 1996.

ladder is a *display* of force, a military demonstration without combat. Finally, the *use* of force is defined as an active military operation against the foreign target.

Do leaders risk higher audience costs when they display or use force? I investigated this question by expanding the set of presidential responses. In one new scenario, the president "sent troops to the region and prepared them for war. The attacking country continued to invade. In the end, the U.S. president did not send our troops into battle, and the attacking country took over its neighbor." In another scenario, the president not only threatened and displayed force, but also "ordered U.S. troops to destroy one of the invader's military bases. U.S. troops destroyed the base, and no Americans died in the operation. The invasion still continued. In the end, the U.S. president did not order more military action, and the attacking country took over its neighbor." The final scenario was identical, except that "20 Americans died in the operation." The new scenarios were administered to a random sample of an additional 1,036 U.S. adults.[18]

Two features made these scenarios appropriate for testing the hypothesis that audience costs increase with the level of hostility. First, the new scenarios differed only in the approach the president took. In all other respects, including background circumstances and the outcome of the crisis, the extra scenarios were identical to each other and to the "stay out/verbal threat" vignettes discussed earlier. Second, the more hostile scenarios nested the less hostile ones: the vignette about the display of force included a threat to use force, and vignettes about the use of force mentioned previous attempts to threaten and display power. Any extra audience costs should, therefore, be due to layering-on higher levels of escalation.

[18] By design, approximately 40 percent of the fresh sample received the "display of force" vignette and the remaining 60 percent were split evenly between the two "use of force" scenarios.

Table 2 summarizes the public reaction associated with each level of escalation. As before, I calculated the percentage of respondents who disapproved either strongly or somewhat when the president escalated and backed down, and subtracted the percentage who disapproved either strongly or somewhat when the president stayed out. This calculation gives the surge in disapproval, or "absolute audience cost," of committing and not following through. Table 2 also presents the relative risk of disapproval, defined as disapproval in the escalation condition divided by disapproval in the stay-out condition.

[TABLE 2 ABOUT HERE]

The estimates in Table 2 show three clear patterns. First, audience costs unambiguously existed in all four scenarios. When the president escalated and did not follow through, disapproval swelled by between 16 and 32 percentage points. Second, audience costs did not increase smoothly with the level of escalation. Based on existing models of audience costs, the president who displayed force should have paid a higher price than the president who merely threatened to use it. In our data, though, the costs were similar: disapproval grew in both scenarios by 16 percentage points with a confidence interval ranging from 10 to 22. The experiment, therefore, provides no evidence that audience costs increase as the president moves from threatening to displaying force. This surprising finding, if replicable, would have significant implications for empirical and theoretical work on military crises.

Third, although audience costs did not rise with each level of escalation, they did exhibit a monotonic trend. The use of arms exposed the president to higher audience costs than either threatening or displaying force, and the loss of lives further raised the price of escalating and then backing down. Each level of audience costs was distinguishable from the previous with probability .95 or better. Overall, the experiments suggested thresholds for the accumulation of audience costs.

5. Are Audience Costs Robust to Variation in International Circumstances?

The evidence thus far confirms that empty commitments cause disapproval to surge. Does this finding hold across a wide range of international contexts? The answer is given in Table 3, which displays audience costs as a function of four standard IR variables: material interests, motive, military power, and political regime.[19] The most striking lesson from Table 3 is that, in every scenario, citizens preferred the president who stayed out to the president who escalated and then backed down. The estimated audience costs were at least 16 points and sometimes as high as 26 points, with confidence intervals that always exceeded zero. Domestic audiences, it seems, disapprove of backtracking against all types of regimes, with varying motivations and military power, whether or not the national interest is at stake.

[TABLE 3 ABOUT HERE]

Although audience costs were always evident, they did vary with the material interests of the escalating state. The price of committing and backing down was smaller—approximately 9 percentage points—when the safety and economy of the United States were not at stake.[20] This difference makes sense. By definition, audience costs depend not only on how the public views empty threats, but also on what the public thinks when the president remains completely aloof. Citizens are naturally more likely to demand military action when they fear for their security and

[19] To increase statistical power I pooled the data from all four levels of escalation, but the main findings hold at each step of the escalation ladder.

[20] The 9-point difference was statistically significant at the .02 level in a one-sided test.

livelihood. It follows that staying out should be less popular in the “hurt” condition than in the “not affect” condition. Moreover, if much of the public disapproves when the president stays out, there may be less potential for disapproval to grow when the president escalates before backing down. Audience costs should, therefore, be smaller when inaction would threaten the national interest.

A similar logic explains why audience costs were somewhat lower in crises against offense-oriented opponents.[21] Previous research found Americans more willing to repel invaders with offensive motives than ones with ambiguous or potentially defensive goals.[22] The same pattern reappeared in our study: dissatisfaction with the “stay out” scenario was more common when the adversary wanted “more power and resources” than when it invaded because of a “longstanding historical feud.” With many citizens already disapproving of the president who remained idle in the face of aggression, there was less potential for audience costs in either absolute or relative terms. Finally, Table 3 shows that audience costs increased with the military power of the adversary, but the opponent’s political regime had little effect.

These findings, though preliminary, suggest that domestic audiences lend more credibility in some international contexts than in others. Threats, for example, may convey more information when issued by leaders who could remain on the sidelines with little risk to their own country. Likewise, threats against status-quo states might be more informative than threats against revisionist ones. Finally, although a thorough analysis of the effects of power would require experiments in many countries, threats by a superpower like the United States may be more revealing when the target is militarily strong than when it is weak.

[21] With a sample of this size, the probability of a difference is at least 8 in 10.

[22] Herrmann, Tetlock, and Visser 1999.

6. The Political Consequences of Backing Down

The experiments described in this article establish a necessary and heretofore unproven condition for audience-cost models of international relations: citizens disapprove of empty threats. Would leaders take this disapproval into account when making foreign policy? The answer surely varies across political systems, but in democracies such as the United States, leaders almost universally view approval as an asset and disapproval as a political cost. Edwards describes the "virtual unanimity" with which presidents, their aides, and participants in the legislative process "assert the importance of the president's public standing" and regard mass approval as "an important source of presidential power."[23]

Foreign policy approval ratings affect the power and incentives of the chief executive in several ways. First, they shape national elections.[24] Citizens consistently list foreign issues among the top problems facing the country, and candidates regularly campaign and speak about foreign policy. This strategy is wise, because judgments about foreign policy are a major component of overall approval, which in turn determines whether sitting leaders can retain office. Aldrich, Sullivan, and Borgida, for example, find that foreign policy issues were just as important as economic ones in predicting how Americans voted in 1980 and 1984, and Gelpi, Reifler, and Feaver offer similarly compelling evidence about the role of foreign policy in the 2004 U.S. election.[25]

---

[23] Edwards 1997, 113-14.

[24] For a literature review, see Aldrich et al. 2006.

[25] Aldrich, Sullivan and Borgida 1989; Gelpi, Reifler and Feaver 2005.

Public approval also enhances the executive's influence over the legislature. In the United States, for example, members of the president's party are more likely to win Congressional elections when the president is popular than when he is not.[26] Moreover, holding constant the partisan composition of Congress, high approval ratings help presidents push initiatives through the legislature and increase the likelihood that vetoes will be sustained.[27] As Krosnick and Kinder explain, "Presidents who are popular in the country tend to have their way in Washington. Popularity is a vital political resource, perhaps the president's single most important base of power."[28] Thus, leaders have good reason to view disapproval as a political liability, and to take care not to damage their standing with the public.

The political fallout from making empty threats would only be magnified if disapproval were concentrated within the most politically active segments of the population. Table 4 shows precisely this pattern.[29] Among respondents who had registered to vote, audience costs averaged 22 percentage points, with a confidence interval from 17 to 27. The analogous effect in the unregistered population was 16 percentage points, which had a wider confidence interval but was still greater than zero with probability .996. Table 4 also distinguishes between nonvoters and active voters: people who recently cast a ballot in a presidential election.[30] Audience costs

[26] Gronke, Koch and Wilson 2003.

[27] For a reconciliation of competing claims in the literature, see Canes-Wrone and de Marchi 2002.

[28] Krosnick and Kinder 1990, 497.

[29] I obtain statistical power by averaging across international circumstances and levels of escalation.

[30] Citizens qualified if they voted in either 2000 or 2004.

among active voters averaged 22 percentage points, versus 15 points for those who had not been to the polls in some time. We can, therefore, be certain that empty threats cause disapproval throughout the population, and especially in the group best positioned to apply electoral penalties.

[TABLE 4 ABOUT HERE]

Empty threats have an even larger effect on citizens who go beyond voting to participate more actively in politics. Following Verba, Schlozman, Brady, and Nie, I classified someone as a political activist if he or she had recently worked for a political campaign, donated money to a campaign, served on a community board, collaborated to solve a community problem, contacted a government official, or attended a political protest or rally.[31] Approximately 29 percent of respondents performed at least one of these activities in the previous twelve months. Audience costs among these activists averaged 34 percentage points, more than double the level among citizens who were not so politically involved. Apparently, the most politically active (and possibly most influential) segments of the population would disapprove at high rates if the executive made threats and did not see them through. These facts bolster the conclusion that backing down would entail domestic political costs.

7. Why Do Citizens Disapprove?

Why, exactly, do citizens disapprove of leaders who escalate crises and then back down? I designed a separate survey of 347 citizens to investigate the micro-mechanisms behind audience costs. As before, citizens considered a situation in which a country invaded its

---

[31] Verba, Schlozman, Brady, and Nie 1993.

neighbor. Some read that the president stayed out; others learned that the president escalated the crisis but did not follow through. In all cases, the attacking country ultimately took over its neighbor. This extra survey went beyond the other experiments, though, by asking citizens to explain the opinions they expressed. After voicing approval or disapproval, participants received a followup: “Could you please type a few sentences telling us why you approve/disapprove of the way the U.S. president handled the situation?” Participants entered their answers directly into a text box, making it possible to analyze each respondent’s account in his or her own words.

For manageability, the study of motivations contained fewer experimental manipulations than the main instrument. In the category of foreign policy strategy, the president either stayed out or displayed force before backing down. The survey also presented a smaller set of background conditions: the invasion would either *hurt* or *not affect* the safety and economy of the United States, but the attacking country was always described as having a strong military, and citizens did not receive information about the motives or political regime of the invader.

This results provided further evidence of audience costs. The president who stayed out received a disapproval score 32 points, while the president who escalated and backed down got negative ratings from 58 percent of the public. The implied cost of 58-32=26 approval points was five times its standard error, and its confidence interval ran from 15 to 35.[32] Thus, the experiment further corroborated one of this article’s main findings, that citizens think more highly of leaders who do nothing than of leaders who commit but do not follow through.

---

[32] This estimate exceeds the value for display of force in Table 2. Why the difference? The text was a bit shorter, so backing down might have appeared starker. Moreover, the adversary in this study always had a strong military, a factor that increases audience costs (see Table 3).

At the same time, the survey provided preliminary evidence about *why* audience costs exist. In the study, 185 citizens considered a scenario in which the president escalated and backed down. Of these, 105 disapproved either strongly or somewhat of the way the president handled the situation. Why did they view the president's behavior negatively? Some did not say, and a few misunderstood the follow-up question or provided an unclassifiable answer, but 87 of the 105 clearly articulated why they had assigned a negative rating.

The 87 open-ended responses fell into three categories. The first category included people who thought the president should have pushed out the invaders, not because the president had made a prior commitment, but simply because it was the right thing to do. Some said the United States had a moral obligation to protect the victims of aggression; others pointed out that the safety and economy of the United States would suffer if the invader took over its neighbor. Fourteen of the 87 participants (approximately 16 percent) answered this way. These citizens probably would have objected as much, and for the same reasons, if the president had stayed out. In fact, most citizens in the control group (in which the president neither threatened nor showed force) justified their disapproval in similar terms. Because these reasons apply equally to all scenarios in which the president let the invasion continue, they cannot be a source of audience costs.

A second group of respondents disliked the fact that the president had escalated in the first place. Some contended that it was not America's responsibility to solve other countries' problems ("I do not feel that the U.S.A. should be the police for the world. We should not have sent troops in this situation.") Others argued that the U.S. government should have focused on its own citizens ("The U.S. has enough problems of our own at this time. We have people that are homeless and hungry. We should take care of our own first"). Roughly 12 percent of respondents offered these dovish or isolationist responses, an often overlooked reason for audience costs.

The vast majority of respondents (72 percent) gave a third reason for disapproving: the president behaved *inconsistently* by saying one thing and doing another. Why did they view inconsistency as problematic? Many complained that waffling would hurt the reputation and credibility of the country. As one citizen explained, “if you say that you are going to do something, you need to do it or else you lose your credibility. It would have been better to ignore the situation completely than to make a public commitment and then not carry it out.” Another respondent wrote: “When a President says something, in this case that he will push back the invading country, he must follow through or lose credibility in the world community. He sent troops and when the threat didn't work, he allowed the invasion to continue. That is a terrible precedent to set.”

A few respondents disliked inconsistency for non-reputational reasons. Two people complained that the president had wasted money by deploying troops but not using them, and eight said the president behaved in a puzzling manner (“Why would he have troops there to help and not do anything to help?”) or had not shown sufficient foresight (“The United States President must not have truly thought things through”). But 61 percent of all disapprovers, and 84 percent of those who complained about inconsistency, denounced the president for breaking his word. By not upholding his commitment to repel the invaders, the president suggested that he and his country could not be trusted.

These responses give preliminary support to a reputation-based theory of audience costs. Early theoretical work proposed that “domestic audiences may provide the strongest incentives for leaders to guard their states' 'international' reputations."[33] Guisinger and Smith extend this insight by developing a model in which reputations for honesty help countries and their agents achieve diplomatic success. Knowing this, citizens will punish leaders for "destroying the country's honest record and thus putting in jeopardy the future benefits of being able to communicate during a crisis."[34] Their model suggests a rational underpinning for audience costs, founded on concerns about reputation.

The evidence in this article is consistent with such a reputational logic. It seems that many citizens value their country's international reputation and disapprove of leaders who mar it. In countries where citizens or elites can hold leaders accountable, the prospect of a domestic backlash should, therefore, create an added incentive to care about international reputations, and thus an extra reason to avoid making empty commitments.

8. Conclusions

This article has offered the first direct analysis of audience costs in a way that avoids problems of strategic selection. The research, based on a set of experiments embedded in public opinion surveys, shows that audience costs exist across a wide range of conditions and increase with the level of escalation. The adverse reaction to empty commitments is evident throughout the population, and especially among politically active citizens who have the greatest potential to shape government policy. Finally, preliminary evidence suggests that audience costs arise from concerns about the international reputation of the country and its leaders.

---

[33] Fearon 1994, 581

34 Guisinger and Smith 2002.

These findings have both substantive and methodological implications for the study of international relations. Substantively, they show how domestic audiences can enhance the credibility of international commitments by punishing leaders who say one thing but do another. This discovery was far from preordained. If citizens had focused on foreign policy outcomes rather than processes, or regarded bluffing as a reasonable strategy, or rewarded leaders for trying before conceding, or cared little about their country's reputation, audience costs would not have emerged. The fact that audience costs arose consistently, across a wide range of conditions, counts as strong evidence that domestic actors can contribute to foreign credibility. Consequently, the article supplies behavioral microfoundations for many leading theories of international security and political economy.

This study also contributes to our understanding of reputation in world affairs. What motivates leaders to protect their international reputations, even at great cost to themselves and others? Domestic audiences heighten the incentive for leaders to care about reputations abroad. Right or wrong, citizens worry that leaders who break commitments will undermine the nation's credibility, and they express strong disapproval when the executive adopts a reputation-damaging strategy. These findings help explain why many leaders strive to protect the national image, and why concerns about reputation shape the way countries behave.

Appendix: Sampling and Interview Methods

The surveys discussed in this article were administered by Knowledge Networks, an internet-based polling firm, with support from the National Science Foundation. By using random digit dialing to recruit participants, and by providing internet access to households that do not have it, Knowledge Networks is able to administer questionnaires to a nationally representative sample of the U.S. population. The surveys took place in July and November 2004, and approximately 76 percent of panelists who were invited to complete the surveys actually did so.

[TABLE A1 ABOUT HERE]

Table A1 compares my sample to the U.S. adult population. National population figures came from the U.S. Census Bureau and the Bureau of Labor Statistics, which provide monthly updates of demographic data through the Current Population Survey (CPS). I computed the benchmarks by pooling data from the July and November 2004 CPS studies (*N*=205,580 adults). The average deviation, in percentage points, between the samples used in this article and the national population was no more than 1.5 percentage points. My sample slightly overrepresents the elderly and residents of the Midwest and the South, while slightly underrepresenting Americans in the highest household income bracket. Even in these categories, though, the deviations are only a few percentage points. Overall, the sample closely matched the population benchmark.[35]

---

35 Was interest in politics higher among respondents than in the nation as a whole? It is hard to know for sure, because the Census Bureau does not collect data on political interest. However, political interest levels in my sample closely matched levels in the General Social Survey (GSS). In my sample, 22 percent of subjects were "very interested" in politics, 40 percent were somewhat interested, 26 percent were slightly interested, and 12 percent were not at all

interested. The comparable GSS figures for 2004 were 21, 49, 20, and 10 percent. In any case, the issue is of minor concern because Table 4 shows large audience costs even among people who do not show much engagement in politics.

References

Aldrich, John H., Christopher Gelpi, Peter Feaver, Jason Reifler, and Kristin Thompson Sharp. 2006. "Foreign Policy and the Electoral Connection." *Annual Review of Political Science* 9: 477-502.

Aldrich, John H., John L. Sullivan, and Eugene Borgida. 1989. "Foreign Affairs and Issue Voting: Do Presidential Candidates Waltz Before a Blind Audience?" *American Political Science Review* 83 (March): 123-141.

Baum, Matthew A. 2004. "Going Private: Public Opinion, Presidential Rhetoric, and the Domestic Politics of Audience Costs in U.S. Foreign Policy Crises." *Journal of Conflict Resolution* 48 (October): 603-31.

Broz, J. Lawrence. 2002. "Political System Transparency and Monetary Commitment Regimes," *International Organization* 56 (Autumn): 861-87.

Brody, Richard A. 1994. "Crisis, War, and Public Opinion: The Media and Public Support for the President." In *Taken by Storm*, eds. W. Lance Bennett and David L. Paletz, 210-27. Chicago: Univ. of Chicago Press.

Canes-Wrone, Brandice and Scott de Marchi. 2002. "Presidential Approval and Legislative Success." *Journal of Politics* 64 (May): 491-509.

Desch, Michael C. 2002. "Democracy and Victory: Why Regime Type Hardly Matters." *International Security* 27 (Fall): 5-47.

Dorussen, Han and Jongryn Mo. 2001. "Ending Sanctions: Audience Costs and Rent-Seeking as Commitment Strategies." *Journal of Conflict Resolution* 45 (August): 395 - 426.

Edwards, George. 1997. "Aligning Tests with Theory: Presidential Approval as a Source of Influence in Congress." *Congress & the Presidency* 24: 113-30.

Eyerman, Joe and Robert A. Hart, Jr. 1996. "An Empirical Test of the Audience Cost Proposition: Democracy Speaks Louder than Words." *Journal of Conflict Resolution* 40 (December): 597-616.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (September): 577-92.

Gaubatz, Kurt Taylor. 1996. "Democratic States and Commitment in International Relations." *International Organization* 50 (Winter): 109-39.

Gelpi, Christopher and Michael Griesdorf. 2001. "Winners or Losers? Democracies in International Crisis, 1918-94." *American Political Science Review* 95 (September): 633-47.

Gelpi, Christopher, Jason Reifler and Peter Feaver. 2005. "Iraq the Vote: Retrospective and Prospective Foreign Policy Judgments on Candidate Choice and Casualty Tolerance." Working Paper, Duke Univ.

Gowa, Joanne. 1999. *Ballots and Bullets: The Elusive Democratic Peace*. Princeton, NJ: Princeton Univ. Press.

Gronke, Paul, Jeffrey Koch, and J. Matthew Wilson. 2003. "Follow the Leader? Presidential Approval, Presidential Support, and Representatives' Electoral Fortunes." *Journal of Politics* 65 (August): 785-808.

Guisinger, Alexandra and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (April): 175-200.

Herrmann, Richard K. and Vaughn P. Shannon. 2001. “Defending International Norms: The Role of Obligation, Material Interest, and Perception in Decision Making.” *International Organization* 55 (Summer): 621-54.

Herrmann, Richard K., Philip E. Tetlock, and Penny S. Visser. 1999. “Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework.” *American Political Science Review* 93 (September): 553-73.

Jensen, Nathan M. 2003. “Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment.” *International Organization* 57 (Summer): 587-616.

Johnson, Valen E. and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15: 163-215.

Krosnick, John A. and Donald R. Kinder. 1990. “Altering the Foundations of Support for the President Through Priming.” *American Political Science Review* 84 (June): 497-512.

Leeds, Brett Ashley. 1999. “Domestic Political Institutions, Credible Commitments, and International Cooperation.” *American Journal of Political Science* 43 (October): 979-1002.

Leventoglu, Bahar and Ahmer Tarar. 2005. “Prenegotiation Public Commitment in Domestic and International Bargaining.” *American Political Science Review* 99 (August): 419-33.

Lipson, Charles. 2003. *Reliable Partners: How Democracies Have Made a Separate Peace*. Princeton, NJ: Princeton Univ. Press.

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

Martin, Lisa L. 1993. "Credibility, Costs, and Institutions: Cooperation on Economic Sanctions." *World Politics* 45 (April): 406-32.

McGillivray, Fiona and Alastair Smith. 2000. "Trust and Cooperation through Agent-Specific Punishments." *International Organization* 54 (Autumn): 809-24.

Partell, Peter J. and Glenn Palmer. 1999. "Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes." *International Studies Quarterly* 43 (June): 389-405.

Prins, Brandon C. 2003. "Institutional Instability and the Credibility of Audience Costs: Political Participation and Interstate Crisis Bargaining, 1816-1992." *Journal of Peace Research* 40: 67-84.

Ramsay, Kristopher W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48 (August): 459-86.

Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53 (Spring): 233-66.

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

Schultz, Kenneth A. 2001b. "Looking for Audience Costs." *Journal of Conflict Resolution* 45 (February): 32-60.

Slantchev, Branislav L. 2006. “Politicians, the Media, and Domestic Audience Costs.” *International Studies Quarterly* 50 (June): 445-77

Smith, Alastair. 1996. “To Intervene or Not to Intervene: A Biased Decision.” *Journal of Conflict Resolution* 40 (March): 16-40.

Smith, Alastair. 1998. “International Crises and Domestic Politics.” *American Political Science Review* 92 (September): 623-638.

Verba, Sidney, Kay Lehman Schlozman, Henry Brady, and Norman H. Nie. 1993. “Citizen Activity: Who Participates? What do they Say?” *American Political Science Review* 87 (June): 303-18.

Walt, Stephen M. 1999. “Rigor or Rigor Mortis? Rational Choice and Security Studies,” *International Security* 23 (Spring): 5-48.

Weeks, Jessica. 2006. “Autocratic Audience Costs: Regime Type and Signaling Resolve.” Stanford University, manuscript.

TABLE 1. *The domestic political cost of making empty threats*

	Public reaction to empty threat	–	*Public reaction to staying out*	=	*Difference in opinion*	*Summary of differences*
Disapprove						
Disapprove very strongly	31 (27 to 36)		20 (17 to 23)		12 (6 to 17)	16 (10 to 22)
Disapprove somewhat	18 (14 to 21)		13 (10 to 16)		5 (1 to 9)	
Neither						
Lean toward disapproving	8 (6 to 11)		9 (7 to 11)		0 (-4 to 3)	-4 (-10 to 2)
Don't lean either way	21 (17 to 25)		21 (18 to 25)		0 (-5 to 4)	
Lean toward approving	8 (6 to 11)		11 (9 to 14)		-3 (-6 to 1)	
Approve						
Approve somewhat	8 (5 to 10)		13 (11 to 16)		-6 (-9 to -2)	-12 (-17 to -8)
Approve very strongly	6 (4 to 9)		13 (10 to 16)		-7 (-10 to -3)	

Note: The table gives the percentage of respondents who expressed each opinion. 95-percent confidence intervals appear in parentheses. Sample size was 477 in the "empty threat" scenario and 650 in the "stay out" scenario. The empty threat scenario involved only a verbal threat; the president did not display or use military force.

TABLE 2. *Domestic audience costs at four levels of escalation*

Level of escalation	*Absolute audience cost*	*Relative risk of disapproval*
Threat of Force	16 (10 to 22)	1.5 (1.3 to 1.7)
Display of Force	16 (10 to 22)	1.5 (1.3 to 1.7)
Use without U.S. casualties	23 (16 to 29)	1.7 (1.5 to 2.0)
Use with U.S. casualties	32 (26 to 39)	2.0 (1.7 to 2.3)

Note: The absolute audience cost is the surge in disapproval, expressed in percentage points. The relative risk is the level of disapproval when the president escalated and backed down, divided by the level of disapproval when the president stayed out. 95-percent confidence intervals appear in parentheses. The sample size was 650 for stay out (the reference category), 477 for threat of force, 420 for display of force, 306 for use without U.S. casualties, and 310 for use with U.S. casualties.

TABLE 3. *Audience costs as a function of international context*

International context	*Disapproval if escalate & BD*	-	*Disapproval if stay out*	=	*Absolute audience cost*
Interests					
Not affect U.S.	51 (48 to 55)		26 (21 to 31)		26 (19 to 31)
Hurt U.S.	56 (52 to 59)		39 (34 to 45)		16 (10 to 23)
Motive					
Historical feud	52 (48 to 55)		29 (24 to 34)		23 (17 to 29)
More power	55 (52 to 59)		36 (31 to 42)		19 (13 to 25)
Regime					
Democracy	53 (50 to 57)		31 (26 to 36)		22 (16 to 28)
Dictatorship	54 (50 to 57)		34 (29 to 40)		19 (13 to 25)
Power					
Strong military	57 (53 to 60)		31 (26 to 36)		26 (20 to 32)
Weak military	50 (47 to 54)		35 (29 to 40)		16 (9 to 22)

Note: The table gives the percentage of respondents who disapproved in each experimental condition. 95-percent confidence intervals appear in parentheses. Sample sizes for the escalation and stay-out conditions, respectively, were 746 and 317 for *not affect the U.S.*; 769 and 313 for *historical feud*; 744 and 337 for *more power*; 740 and 328 for *democracy*, and 733 and 322 for *dictatorship*; 767 and 333 for *would hurt the U.S.*; 728 and 330 for *strong military*; and 785 and 320 for *weak military*.

TABLE 4. *Audience costs by level of political participation*

Level of participation	*Disapproval if escalate & BD*	-	*Disapproval if stay out*	=	*Absolute audience cost*
Registered	56 (53 to 59)		34 (30 to 38)		22 (17 to 27)
Not registered	42 (36 to 49)		26 (17 to 37)		16 (4 to 27)
Voter	57 (54 to 60)		34 (30 to 39)		22 (17 to 28)
Non-voter	44 (38 to 50)		29 (21 to 37)		15 (5 to 24)
Activist	63 (58 to 68)		29 (23 to 36)		34 (26 to 42)
Non-activist	50 (47 to 53)		34 (30 to 39)		16 (11 to 21)

Note: The table gives the percentage of respondents who disapproved in each scenario. 95-percent confidence intervals appear in parentheses. Sample sizes for the escalation and stay-out conditions, respectively, were 1120 and 476 for citizens who were *registered* to vote, 227 and 88 for citizens who were *not registered*, 1026 and 432 for *voters*, 302 and 126 for *non-voters*, 427 and 190 for *activists*, 1031 and 440 for *non-activists*.

TABLE A1. *Characteristics of the National Population and the Survey Sample (%s)*

		Adult U.S. Population	Audience Costs Sample	Absolute Deviation
Gender	Male	47.7	47.9	0.2
	Female	52.3	52.2	0.2
	Average Deviation			0.2
Age	18-24	11.8	10.4	1.4
	25-34	17.1	15.3	1.9
	35-44	20.0	20.0	0.0
	45-54	19.9	19.7	0.2
	55-64	14.3	15.2	0.9
	65 or older	16.9	19.6	2.6
	Average Deviation			1.2
Education	No high school diploma	14.1	14.3	0.3
	High school diploma	32.7	30.7	2.0
	Some college	19.2	23.8	4.7
	Associate Degree	8.5	7.5	0.9
	College degree	25.6	23.6	2.0
	Average Deviation			2.0
Income	Less than $10,000	7.1	7.2	0.1
	$10,000 - 24,999	17.2	18.7	1.5
	$25,000 - 49,999	28.9	34.2	5.2
	$50,000 - 74,999	20.5	21.4	0.9
	$75,000 or more	26.3	18.6	7.7
	Average Deviation			3.1
Marriage	Married	58.4	58.3	0.1
	Not married	41.6	41.8	0.1
	Average Deviation			0.1
Race	White	84.0	84.7	0.8
	Non-white	16.0	15.3	0.8
	Average Deviation			0.8
Region	Midwest	24.3	26.3	2.1
	Northeast	21.4	18.3	3.1
	South	30.1	34.2	4.1
	West	24.3	21.2	3.1
	Average Deviation			3.1
Total Average Deviation				1.5

The Effects of International Agreements on Foreign Policy Preferences

Michael Tomz
Stanford University
tomz@stanford.edu

January 2007

Preliminary. Comments welcome!

I am grateful for financial support from the National Science Foundation (CAREER grant SES-0548285). The staff at Knowledge Networks—especially Mike Dennis and Sergei Rodkin—provided invaluable assistance with this project.

Abstract: Do international agreements transform the way citizens think about foreign policy, or are such agreements mere scraps of paper that have little effect on policy preferences? I investigate this question by offering the first-ever experimental analysis of international agreements. The experiments, embedded in public opinion surveys, directly measure the effect of international agreements on foreign policy preferences and, at the same time, avoid problems of endogeneity that have stymied previous research. Citizens, I find, are far more likely to oppose policies that would violate international legal agreements than to oppose *otherwise identical* policies that would not trammel upon existing pacts. Moreover, the political effects of international agreements diffuse widely throughout the population, affecting conservatives as well as liberals and crossing other demographic divides. But citizens do not follow international legal agreements blindly. When the case for pursuing a policy is strong enough, a clear majority will endorse the policy, even though it would violate international law. Thus, international legal agreements strongly influence—but do not dictate—preferences about foreign policy.

Under what conditions do international legal agreements affect the behavior of states? Few questions are as central to the study and practice of global politics. According to the United Nations Treaty Series, more than 50,000 international agreements are now in force. Although these agreements are "legally binding," sovereign states have no higher power to which they can appeal when another party reneges. The tension between legal theory and international anarchy has stimulated much debate about whether international law is irrelevant, or whether it truly influences countries' policies.

Previous research has not settled the debate. As Simmons (1998: 89) points out, "Several studies have tried to demonstrate a correlation between legal standards and state behavior, sometimes employing large databases and statistical techniques, but most are unconvincing in demonstrating causation, or even in providing an explanatory link between the actions taken and the existence of agreements or normative considerations." This ambiguity is not necessarily the fault of political scientists or lawyers; rather, it stems from the challenging nature of the subject itself. As most contributors to the literature now recognize, problems of endogeneity make it extremely difficult to know whether, and in what ways, international agreements influence state behavior.

In this paper, I advance our understanding of international agreements by supplementing the analysis of observational data with experiments involving randomized treatment and control. A pure field experiment, in which the researcher forces some governments to sign international agreements and others to abstain, is clearly out of the question, but other interventions are both feasible and informative. I summarize the outcome of a survey-based experiment that directly measured the effect of international agreements and, at the same time, avoided problems of endogeneity that have stymied previous research.

Data from the experiment support three conclusions. First, international legal agreements transform policy preferences. Citizens are far more likely to oppose policies that would violate international law than to oppose *otherwise identical* policies that would not trammel upon existing treaties and other agreements. Second, the political effects of international law diffuse widely throughout the population, affecting conservatives as well as liberals and crossing other demographic divides. Third, citizens do not follow international law blindly. When the case for pursuing a policy is strong enough, a clear majority will endorse the policy, even though it would violate international law. In summary, international law shapes—but does not dictate—the foreign policy preferences of citizens.

In the remainder of this paper, I discuss the limits of existing evidence, present the findings from a survey-based experiment, and conclude by enumerating the strengths and weaknesses of my experimental approach and discussing avenues for future research.

1. The Limits of Existing Evidence

International relations scholars disagree about the effect of formal international agreements. Some say that such agreements have little independent effect on foreign policy (e.g. Mearsheimer 1994, 2001). In a world of sovereign states, participation in treaties and other legal pacts is entirely voluntary. Countries sign agreements they are predisposed to follow, and they remain parties while the agreement serves their interests. If the commitment ever becomes inconvenient, however, countries may withdraw unilaterally. In fact, most treaties allow parties to end their legal obligations by giving just a few months' advance notice. When a graceful exit is not possible, parties can break the agreement without fear of punishment by a world government.

Other analysts argue that formal agreements shape the way countries behave (e.g. Keohane 1984; Fortna 2003).[1] According to this view, when leaders sign an international agreement, it becomes more costly to take actions the agreement forbids and less costly to pursue policies the agreement condones. Why might the agreement change incentives? One plausible mechanism is reputation: putting an agreement in writing may increase the reputational cost of reneging. A second mechanism concerns norms, rather than interests: perhaps people feel it would be morally wrong to break promises to foreigners. Under either of these two mechanisms, international agreements would shape the way countries behave.

Due to limitations of existing data, it has been extremely difficult to resolve the debate. We simply do not know the conditions under which international agreements matter, or when the act of signing an international agreement is most likely to demonstrate credibility.

Previous research has relied entirely on the historical record to estimate the effect of international agreements. Have countries that signed environmental protection treaties polluted less than countries that did not sign? Has respect for human rights, arms control, and free trade been greater among countries that entered agreements on these topics than among countries that did not?

If agreements arose from a purely random process, the use of historical data would be unproblematic. In reality, though, countries choose whether and on what terms to enter international agreements. Thus, it is hard to know whether the historical correlation between agreements and behavior is a consequence of the agreement itself, or is due to cross-national differences in the baseline propensity to take the kinds of actions the agreement requires. Most countries honor their agreements most of the time, but this does not prove that agreements shape

[1] For an excellent review of the literature as applied to international trade and the GATT/WTO, see Busch and Reinhardt (2002).

foreign policy. According to skeptics, agreements reflect but do not change the pre-existing interests of states.

We can bring the existing debate—and the roadblock of endogeneity—into sharper relief by drawing on Rubin's (1974) counterfactual account of causality. Suppose we are interested in Y_i, a measure of country *i*'s behavior on a given issue at a particular time. The impact of an international agreement on Y_i is $\delta_i = Y_{i1} - Y_{i0}$, where Y_{i1} represents the way *i* would behave if party to the agreement, and Y_{i0} signifies how the same country would behave if not party to the agreement. The quantity δ_i tells what difference, if any, the agreement makes.

Unfortunately, the causal effect δ_i is unobservable. We can imagine how country *i* might behave in both the agreement and the no-agreement conditions, but we cannot observe both Y_{i1} and Y_{i0} for the same *i* at the same time. After all, no country can be in the treatment and control regimes simultaneously. Holland (1986) has called the inability to observe the same unit under both treatment and control "the fundamental problem of causal inference."

Although the individual-level effect δ_i is beyond reach, scholars have tried to infer *the average* causal effect δ by comparing the observed *Y* for countries that signed the agreement with the observed *Y* for countries that did not. In this context, the average effect of the agreement is estimated as $\hat{\delta} = (Y_1 \mid A = 1) - (Y_0 \mid A = 0)$, where the indicator *A* takes a value of 1 when the country is party to the agreement and 0 otherwise.

This standard estimator is equal to the true effect δ plus two potential sources of bias. The first source of bias, "different baseline propensities," arises when signatories and non-signatories differ in their fundamental tendency to do *Y*, even in the absence of an international agreement. The second source of bias, "different treatment effects," arises when the agreement would produce stronger (or weaker) effects on the group that actually signed than on the group that did not.

With a bit of algebra, we can show that

$$\hat{\delta} = \delta + \left[(Y_0 \mid A=1) - (Y_0 \mid A=0)\right] + \pi\left[(\delta \mid A=1) - (\delta \mid A=0)\right],$$

where π is the proportion of the sample that did not join the agreement, $\delta \mid A=1$ is the average effect on those who actually signed, and $\delta \mid A=0$ is the average effect the agreement would have exerted on the remaining countries, if contrary to fact they had chosen to sign. The first term in brackets gives the bias from different baseline propensities; the second gives the bias from heterogeneous treatment effects.

Expressing $\hat{\delta}$ in this way helps clarify the claims and the limitations of the existing literature. Researchers have found that countries often comply with international agreements, and in some cases the observed level of *Y* differs systematically between signatories and non-signatories. To the skeptic, these estimated effects are artifacts of different baselines: those who signed were more inclined to do *Y* in the first place.

Essentially, skeptics argue that that δ is approximately zero but our estimate $\hat{\delta} \neq 0$ because the conditional mean (Y_0|A=1) exceeds the conditional mean (Y_0|A=0). If we could eliminate baseline differences between the two groups, the skeptic argues, the apparent effect of the international agreement would disappear. To convince the skeptic that international agreements matter, and to obtain unbiased estimates of the causal effect more generally, it is important to remove any baseline differences. I argue below that experiments can achieve this goal by design, whereas observational studies can do so only with great difficulty.

Against the skeptics, legalists argue that a country can alter its interests and behavior by signing an international agreement. Moreover, sophisticated legalists hypothesize that agreements affect different countries to different degrees. For example, Hathaway (2005) and Raustiala and Slaughter (2002) suggest that the relationship between international law and state behavior depends on domestic institutions such as the judiciary, the media, political parties, and interest groups. Others claim that sensitivity to agreements varies with the rule of law or the degree of democracy. These and other domestic institutions vary considerably across states. As a consequence, δ_i should differ from one *i* to the next.

Heterogeneity in δ_i creates a second source of bias: a correlation between signatories and susceptibility to treatment. If leaders are rational, they will weigh the anticipated effects of the agreement when deciding whether to sign. This rational behavior introduces a systematic relationship between signatory status (A_i) and the treatment effect (δ_i). Using the previous notation, $(\delta \mid A=1) \neq (\delta \mid A=0)$ in the presence of self-selection. This heterogeneity is, of course, interesting in its own right and an important subject of study. Unless the heterogeneity is controlled, however, $\hat{\delta}$ will be a biased estimate of the treatment effect.

To draw valid inferences from non-experimental data, we need statistical correctives that allow us to approximate the attributes of a genuine experiment. Some researchers address this problem with control variables: they model foreign policy as a function of international agreements and controls that correct for differences between signatories and non-signatories (e.g. Simmons 2000). The goal is to make signatories and non-signatories comparable after conditioning on the *X*'s, such that any systematic difference in behavior would reflect the causal effect of the agreement, rather than distinct baselines or different sensitivities to treatment.[2]

[2] Heckman-type selection models can be viewed variants of the same strategy: using a function of one or more control variables to address the problem of endogeneity. For applications of this approach to IMF agreements, see Przeworski and Vreeland (2000) and Vreeland (2003).

The success of this approach depends on a comprehensive set of controls, however. To solve the bias problem with control variables, the researcher must condition on all variables that correlate with the outcome and membership in the agreement (Besley and Case 2000). This can be quite a challenge, made more severe by informational asymmetries in international relations. Governments have private information, which they withhold not only from other countries but (presumably) from academic researchers, as well! When governments have pertinent but private information about their baseline interests or their sensitivity to treatment, the set of control variables is likely to be incomplete and estimates of the agreement's effect will be biased.

Some of these problems can be minimized through the use of panel data. If countries are tracked over a number of years, the insertion of fixed effects for countries or dyads can help correct for unobserved heterogeneity. Hathaway (2002), Goldstein, Rivers and Tomz (2007), Simmons (2004), and Tomz, Goldstein, and Rivers (2007) have used this approach to study the effects of international agreements on human rights and international trade. For the strategy to succeed completely, though, the determinants of state policy must be additive and time-invariant. If the decision to enter an agreement and the choice of Y depend on common variables that change over time, omitting those variables will lead to biased estimates of δ, even in the presence of fixed effects (Besley and Case 2000).

There is a second statistical option. Instead of using controls that are correlated with both the treatment and the outcome, one could seek an instrument that affects the outcome only indirectly, via the treatment variable. With a technique such as two-stage least squares regression, the instrument can be used to obtain consistent estimates of the treatment effect. This approach makes sense in theory but has serious problems in practice: it has been nearly impossible to find valid instruments: ones that correlate strongly with the presence or absence of an international agreement but have no independent bearing on foreign policy.

In summary, problems of endogeneity make it difficult to infer the effects of international agreements. Several studies have tried to address this problem via control variables, but many questions about causality remain. The next section explains how experiments can isolate the causal effects of international agreements.

2. An Experiment-Based Approach

The core idea of this paper is to supplement observational studies by embedding experiments in interviews with citizens and elites. Some interviewees hear about a hypothetical or historical foreign policy situation in which leaders have signed a legally binding international agreement. Others consider exactly the same situation, *sans* any international agreement. By comparing the views of participants in the treatment condition (international agreement) versus the control condition (no agreement), we can isolate the effect of international agreements on policy preferences and beliefs.

As a first step toward implementing this approach, I designed an experiment about international trade agreements. The experiment aimed to quantify the effect of trade pacts *in the context* of many other foreign policy considerations, such as humanitarianism and the national interest. This represents a relatively hard test. If trade agreements matter even in the face of competing or redundant concerns, they are likely to be influential in less stringent settings.

The experiment, which was administered over the internet to a nationally representative sample of 1000 U.S. adults in July 2005,[3] began as follows: "The next question is about foreign

[3] The experiment was fielded by Knowledge Networks, and interviews were conducted over Web TV and the internet.

policy. Some leaders want the United States to prohibit trade with the country of Burma. They say we should neither buy products from Burma nor sell products to Burma. Experts who have studied this proposal agree on several points. Please consider each point carefully, and then tell us what you think."

I presented each respondent with two or more of the following points.

US economy: "The proposal would help the U.S. economy. Many Americans are getting laid off because of competition from Burma. If we stop trading with Burma, there will be more jobs and higher wages in the United States."

Human rights: "The proposal would help human rights. In Burma, the government kills political opponents and does not allow free speech. By stopping trade with Burma, we can pressure the government to start respecting basic rights."

Burmese economy: "The proposal would hurt the Burmese economy. Burma sells $300 million in products to the United States each year. If we stop trading with Burma, people in that country will lose their jobs, and poverty will rise."

International law: "The proposal would violate international law. The United States has signed treaties that make it illegal to limit trade with Burma. If we stop trading with Burma, we will be breaking international law."

Neutral argument: "The proposal would change our trade relations. The United States trades with many countries. If we stop trading with Burma, we will no longer suffer the costs (if any) nor will we get the benefits (if any) of trade with that particular country."

I randomly assigned each respondent to one of nine groups, each of which confronted a different configuration of arguments (see Table 1). For example, respondents in group 1 learned that the proposal to cut off trade would help the US economy but hurt the Burmese economy. Participants in group 2 received exactly the same considerations, plus they were told that the plan would violate international law by contravening trade agreements the United States had signed. This between-subject design makes it possible to measure the effect of international agreements by comparing the preferences of citizens in group 1 versus group 2. The third group received a "neutral" argument, which I included to quantify the influence of international law while holding the number of arguments constant (e.g. compare group 2 versus group 3).

[TABLE 1 ABOUT HERE]

The remaining entries in Table 1 altered the mix of "pro" arguments. Conditions 4-6 contained no mention of the US economy, but they stressed that the plan would help human rights by putting pressure on the Burmese economy. Finally, I posed a double-challenge to trade agreements by constructing scenarios in which the decision to eliminate trade would both serve US economic interests and advance human rights (groups 7-9).[4] To guard against any order effects, I randomized not only the group assignments but also the sequence of arguments within

[4] By design, all respondents heard that the plan would hurt the Burmese economy. I included the factor to balance and fill-out the list of considerations.

each group. For instance, half the participants in group 1 received the argument about the US economy before the argument about the Burmese economy; the other half considered the opposite succession of points.

After presenting these arguments, I asked citizens to express their foreign policy preferences. "How good or bad an idea is it for the United States to prohibit trade with Burma?" The response options were: extremely good, moderately good, slightly good, neither good nor bad, slightly bad, moderately bad, and extremely bad. By analyzing the responses, one can see how the presence or absence of an international legal agreement affected policy preferences.

Before computing the effect of such agreements, it seemed prudent to confirm that the treatment and control groups were balanced on baseline covariates that could affect foreign policy preferences. I estimated a logistic regression in which the dependent variable was the dichotomous treatment (international law mentioned=1, not mentioned=0) and asked whether any demographic or contextual variables predicted membership in the treatment group. Not one of the many variables in the model—ideology, party identification, gender, age, race, education, income, and a variety of other demographic factors—had a statistically significant effect on the probability of being in the treatment group. Based on a likelihood ratio test, we cannot reject the hypothesis that the relationship between the treatment and *all* baseline variables was zero.[5]

Having established that the treatment was random, I proceeded to estimate the effect of international agreements. Of respondents who took a side (all respondents except those who answered "neither good nor bad"), I computed the share who thought it would be a bad idea to sever commercial relations with Burma. To the extent that international agreements matter, this

[5] The likelihood ratio test statistic, 22.26, was distributed chi-squared with 23 degrees of freedom. If all coefficients were zero, we would observe a test statistic that large roughly half the time.

share should be significantly higher when international law is mentioned (groups 3, 6, and 9) than when it is not.

In all cases, I obtained point estimates and confidence intervals via Bayesian simulation. Specifically, I modeled the proportion of people who opposed the policy proposal as a beta distribution with a uniform prior (see Johnson and Albert 1999, 11). When an international agreement exists, this proportion π_1|data is distributed as beta(b_1+1, $N_1 - b_1$ +1), where the subscript 1 indicates the treatment regime, N_1 is the number of respondents who received the treatment, and b_1 is the number of people who thought it would be bad to cut trade with Burma.

Without an international agreement, the proportion of naysayers is π_0|data and distributed as beta(b_0+1, $N_0 - b_0$ +1), where the subscript 0 signifies the control regime. By drawing random variates from these independent beta distributions, we can obtain the full posterior distribution (and therefore point estimates and confidence intervals) of $\delta = \pi_1 - \pi_0$, the effect of the international agreement. I employ this approach in the remainder of the paper. Other statistical methods, including ordered probit analysis and comparison of means using the full seven-point scale from extremely bad to extremely good, lead to the same conclusions.

Table 2 shows that international legal agreements powerfully affect the preferences of citizens. When no international agreement was mentioned (conditions 1, 3, 4, 6, 7, and 9), approximately 27 percent of respondents who took a side deemed it bad to prohibit trade with Burma. This percentage jumped 17 points when respondents were told that the policy would violate international law. The 95 percent confidence interval around this effect ranged from 10 to 25, so we can be quite sure that the shift in policy preferences did not arise by chance alone. The data in Table 2 thus provide strong microfoundations for the view that international legal agreements do not simply reflect, but can actually change, preferences about foreign policy.

[TABLE 2 ABOUT HERE]

Table 3 divides the sample into three parts, based on non-legal considerations that respondents weighed. The table supports three inferences. First, on matters of international trade, economic arguments tend to sway citizens more effectively than appeals to human rights. When experts concluded that the trade barriers would improve human rights, 37 percent of respondents who took a side nonetheless opposed the measure (Table 3, column 1). Citizens who heard about the U.S. economy, rather than human rights, were significantly more likely to support the proposal.

Second, appeals to human rights and economic interest reinforce each other. Citizens do not regard these considerations as equivalent, and they are more willing to support a foreign policy that would serve both objectives than a policy that would enhance only one. In fact, only 15 percent of citizens who expressed an opinion actually disapproved when told that we could create U.S. jobs and improve human rights by eliminating trade with Burma.

[TABLE 3 ABOUT HERE]

Third, the table confirms that international law can change foreign policy preferences, even in the face of powerful counterarguments. Consider the first row of Table 3, which pertains to people who heard that the policy initiative would help human rights. For those citizens, the mere insertion of an international agreement converted naysayers from a minority (37 percent) to a majority (54 percent). The estimated effect of international law in this situation was 17 percentage points, with an unambiguously positive confidence interval. The second row of the table summarizes opinion among citizens who heard that the proposal would help the U.S.

economy. Here, international law exerted a smaller but still discernable effect, leading to a 10-percentage-point swing in policy preferences. This estimate, though somewhat imprecise, was nonetheless greater than zero with a Bayesian p-value of 0.06.

Surprisingly, international law remained potent even when counter-arguments were extremely strong (Table 3, row 3). The proposal to create jobs and improve human rights was highly popular among members of the control group, but the same proposal elicited scorn from 37 percent of citizens in the treatment condition. Under this scenario, the presence of an international agreement more than doubled the share of citizens who opposed the foreign policy. The overall effect was 22 points, with a 95-percent confidence interval from 9 to 35. Thus, contrary to the skeptics, it appears that international legal agreements can affect preferences, even when the decision to follow international law might not serve either economic self-interest or humanitarian concerns.

Further analysis suggests that the political effects of international law diffuse widely throughout the population, affecting conservatives as well as liberals and crossing other demographic divides. Table 4 presents a breakdown of estimates by demographic group. Regardless of political ideology, party identification, gender, education or income, international law substantially changes preferences about foreign policy. The effect sizes in the table range from 12 to 25 percentage points, and all are distinguishable from zero with a Bayesian p-value of .001 or better. Moreover, although some demographic groups appear more sensitive to international law than other groups, we cannot affirm these differences with a high level of confidence. Overall, we cannot reject the null hypothesis that international law has similar effects on all demographic groups.

3. Conclusions

In at least two ways, the experiments in this paper shed new light on the effects of international legal agreements. First, the experiments overcame problems of endogeneity that have hampered previous studies. In the experiments, I assigned treatment and control randomly without reference to background features of the situation or the respondent. As a result, there was no significant correlation between the agreement and baseline propensities or sensitivity to treatment. This greatly simplified the problem of inference: I obtained unbiased estimates of the agreement's effect through tabular analysis of the experimental data. Randomization eliminated the need for scores of regressions with control variables, which must be used in observational studies to balance the treatment and control groups.

Second, the experiments helped strengthen the microfoundations of international legal theory by revealing how legal agreements affect the preferences of individuals. Researchers have found correlations between international agreements and policy outcomes. Some contend that the relationship is spurious, whereas others argue that international agreements change the cost/benefit calculations of citizens. The evidence in this paper supports the second interpretation. At an individual level, citizens are more reluctant to pursue policies that would violate international law than to pursue otherwise identical policies that are not enshrined in a legal commitment.

The experimental method in this paper could be extended to answer a wide range of questions about international agreements. The prevailing research strategy, which relies entirely on observational data, is essentially passive. To study a particular variable, researchers must wait for natural processes to generate the variation they need, in quantities large enough to support statistical analysis or in patterns convenient enough to permit controlled case studies. A passive strategy has significant limitations. Some factors may exhibit minimal variation or be highly collinear with other factors, and some values may occur too rarely to support precise estimates. An experimental approach can overcome these limitations by allowing full control over the explanatory variables.

In particular, subsequent studies could vary not only the presence but also the form of the international agreement, thereby shedding light on the effects of institutional design (Koremenos, Lipson and Snidal 2001). Lipson (1991) hypothesizes that the costs of reneging increase with the precision of the agreement, the formality by which it was conveyed, and level of government that authorized it. Rosendorff and Milner (2001) add that the penalty for deviating from commitments can be lower in the presence of escape clauses. Finally, my work on international debt (Tomz 2007) shows that lenders will excuse defaults that occur because of a fundamental change in circumstances or widespread noncompliance by other parties. With experiments, we can test whether citizens and elites take similar contingencies into account when thinking about treaties and other international agreements.

Of course, the experimental approach is not infallible. Indeed, experiments are vulnerable on precisely the dimension where observational data is most compelling: external validity. Voters and elites might behave differently in an interview than in real foreign policy situations. Differences could emerge because respondents know they are subjects of a study, because the interviewer can offer only limited background information, or because emotion plays a different role in interviews than in actual politics. It is not clear how serious these problems would be, but it is important to acknowledge them.

To some extent, concerns about external validity can be minimized by making the scenario as convincing as possible and replicating the experiments with different question wording and sample frames, to increase confidence in the generality of the results. Ultimately, though, the evidence from experiments should be combined with observational data to obtain a fuller understanding of international agreements. Every methodology has its limitations. The best way to make progress on complicated topics is to analyze data from multiple sources. Thus, the evidence in this paper complements a growing body of high-quality research that others have done with historical data.

Works Cited

Besley, Timothy and Anne Case. 2000. "Unnatural Experiments? Estimating the Incidence of Exogenous Policies." *Economic Journal* 110 (November): F672-94.

Busch, Marc L. and Eric Reinhardt. 2002. "Testing International Trade Law: Empirical Studies of GATT/WTO Dispute Settlement." In Daniel L. M. Kennedy and James D. Southwick, eds., *The Political Economy of International Trade Law* (New York: Cambridge Univ. Press, 2002): 457-81.

Fortna, Virginia Page. 2003. "Scraps of Paper? Agreements and the Durability of Peace," *International Organization* 57:337-72.

Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61: 37-67.

Hathaway, Oona A. 2002. "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111: 1935.

Hathaway, Oona A. 2005. "Between Power and Principle: An Integrated Theory of International Law." *University of Chicago Law Review* 72: 469-536.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81, no. 396 (December): 945-60.

Johnson, Valen E. and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Keohane, Robert. 1984. *After Hegemony*. Princeton, NJ: Princeton Univ. Press.

Koromenos, Barbara, Charles Lipson and Duncan Snidal. 2001. "The Rational Design of International Institutions," *International Organization* 55, no. 4 (Autumn): 761-99.

Lipson, Charles. 1991. "Why are Some International Agreements Informal?" *International Organization* 45, no. 4 (Autumn): 495-538.

Mearsheimer, John J. 1994. "The False Promise of International Institutions," *International Security* 19:5-49.

Mearsheimer, John J. 2001. *The Tragedy of Great Power Politics*. New York: Norton.

Przeworski, Adam and James Raymond Vreeland. 2000. "The Effect of IMF Programs on Economic Growth." *Journal of Development Economics* 62: 385-421.

Raustiala, Kal and Anne-Marie Slaughter. 2002. "International Law, International Relations and Compliance." *Handbook of International Relations*, pp. 538-58. NY: Sage.

Rosendorff, B. Peter and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55, no. 4 (Autumn): 829-57.

Rubin, Donald. 1974. "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies." *Journal of Educational Psychology* 66: 688-701.

Simmons, Beth. 1998. "Compliance with International Agreements." *Annual Review of Political Science* 1:75-93.

Simmons, Beth. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94, no. 4 (December): 819-35.

Simmons, Beth. 2004. "International Law Compliance and Human Rights." Working Paper, Harvard University.

Tomz, Michael, Judith Goldstein and Doug Rivers. 2007. "Membership Has Its Privileges: Understanding the Effects of the GATT and the WTO on World Trade." Forthcoming, *American Economic Review*.

Tomz, Michael. 2007. *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton, NJ: Princeton Univ. Press.

Vreeland, James. 2003. *The IMF and Economic Development*. New York: Cambridge University Press.

Table 1: Experimental Conditions

Each respondent was assigned to one of the following groups and presented with the arguments marked by X's.

Group	U.S. economy	Human rights	Burmese economy	International agreement	Neutral argument	Sample size
1	X		X			128
2	X		X	X		114
3	X		X		X	125
4		X	X			114
5		X	X	X		103
6		X	X		X	99
7	X	X	X			113
8	X	X	X	X		88
9	X	X	X		X	116

Table 2: The Effect of International Law

Table gives the percent of respondents who opposed cutting trade with Burma
Bayesian 95% confidence intervals appear in parentheses.

Violates International Law?		
No	Yes	Effect
27 (23 to 31)	44 (38 to 51)	17 (10 to 25)

Table 3: Effect of International Law, Conditional on Counter-Arguments
Table gives the percent of respondents who opposed cutting trade with Burma
Bayesian 95% confidence intervals appear in parentheses.

	Violates International Law?		
	No	Yes	Effect
Human rights only	37 (30 to 45)	54 (43 to 65)	17 (4 to 30)
U.S. economy only	30 (23 to 37)	40 (30 to 51)	10 (-3 to 23)
Both arguments	15 (10 to 21)	37 (26 to 49)	22 (9 to 35)

Table 4: The Effect Cuts Across Demographic Divisions

Table gives the percent of respondents who opposed cutting trade with Burma
Bayesian 95% confidence intervals appear in parentheses.

	Violates International Law?			
	No	Yes	Effect	Difference
Liberals	28 (21 to 36)	53 (41 to 65)	25 (11 to 40)	10 (-10 to 29)
Conservatives	31 (25 to 38)	47 (35 to 58)	15 (2 to 29)	
Democrats	27 (21 to 33)	48 (37 to 59)	21 (9 to 34)	4 (-14 to 22)
Republicans	25 (19 to 32)	43 (32 to 53)	17 (5 to 30)	
Females	26 (21 to 31)	45 (37 to 54)	19 (9 to 30)	-4 (-11 to 20)
Males	28 (22 to 34)	43 (33 to 53)	15 (4 to 26)	
Some College	28 (23 to 33)	47 (38 to 56)	19 (9 to 29)	5 (-11 to 20)
No College	26 (20 to 32)	40 (31 to 50)	15 (3 to 26)	
High Income	23 (18 to 28)	46 (36 to 56)	24 (12 to 35)	12 (-4 to 27)
Low Income	31 (25 to 36)	43 (34 to 51)	12 (2 to 22)	