

INTERNATIONAL DIMENSIONS OF AMERICAN PUBLIC OPINION

Methodological Critiques and Three Empirical Essays

BAOBAO ZHANG

Advisors:

Alan Gerber (chair), Allan Dafoe, Devin Caughey, Nikolay Marinov
Political Science Department
Yale University

April 19, 2013 – Final Draft

It is as particular entities that states enter into relations with one another. Hence their relations are on the largest scale a maelstrom of external contingency and the inner particularity of passions, private interests and selfish ends, abilities and virtues, vices, force, and wrong. All these whirl together, and in their vortex the ethical whole itself, the autonomy of the state, is exposed to contingency. The principles of the national minds are wholly restricted on account of their particularity, for it is in this particularity that, as existent individuals, they have their objective actuality and their self-consciousness. Their deeds and destinies in their reciprocal relations to one another are the dialectic of the finitude of these minds, and out of it arises the universal mind, the mind of the world, free from all restriction, producing itself as that which exercises its right – and its right is the highest right of all – over these finite minds in the 'history of the world which is the world's court of judgement.'

— Georg Wilhelm Friedrich Hegel, *The Philosophy of Right*

ABSTRACT

For the past decade, survey experiments have become the method *du jour* for studying public opinion. In the field of international relations, researchers have tested theories and findings from observational studies using survey experiments. Building upon this small but growing literature, I conduct three sets of survey experiments to understand Americans' attitudes about foreign policy.

An important substantive contribution of this senior essay is that I demonstrate that some American voters' attitudes towards foreign policy are influenced by international elites. The traditional view of American politics suggests that voters take cues mainly from domestic elites. Nevertheless, my experimental data show that the UN Security Council and foreign countries can shape some Americans' support for war.

On the methodological front, I tackle the problem of confounding vignettes in survey experiments. Poorly designed vignettes cause confounding because they change subjects' beliefs about factors other than the one researchers want to change. As a result, these vignettes do not operationalize the construct researchers seek to test. I detect confounding through placebo tests and improve experiment design by introducing more precise vignettes.

ACKNOWLEDGEMENTS

It takes a village to write a senior essay at Yale. And I thank the village that is the Yale community. First, I would like to acknowledge the Yale College Council of Masters for the Richter Summer Research Fund, the Institution for Social and Policy Studies, Calhoun College, and Allan Dafoe for their financial support. I am also grateful for the institutional support of the Yale political science department.

Second, I thank my advisors Alan Gerber, Allan Dafoe, Devin Caughey, and Nikolay Marinov. You are better group of advisors than most PhD dissertation committees. (Although some of you might serve on my dissertation committee.) Alan Gerber has been instrumental in teaching me experimental methods and good research design. I am very fortunate to have found Allan Dafoe, who provided me with a crucial idea for this senior thesis. His innovative thinking and persistent guidance makes him an excellent co-author and mentor. Although he is a faculty member at the Massachusetts Institute of Technology, Devin Caughey is kind enough to become an advisor. His helpful comments and clarity of exposition has improved my writing — every political scientist needs the voice of a historian. Lastly, I acknowledge the support of my long-time mentor Nikolay Marinov, who has guided me through this research process long before my senior year. He first introduced me to the technological skills and international relations theory that forms the backbone to this project.

Other Yale faculty I want to thank include John Bullock, who helped me design numerous surveys and formalize key concepts in this thesis. I also like to thank Daniel Butler for telling me truthfully that some of my ideas are crazy. Outside of Yale, I would like to thank Donald Green (Columbia), my mentor and the man who first introduced me to experimental methods. In addition, I am grateful for comments from Dustin Tingley (Harvard), Jas Sekhon (UC Berkeley), Michael Tomz (Stanford), and Eleonora Mattiacci (Ohio State). Finally, where would I be without the “anxiety” inducing influence of David Broockman (UC Berkeley), who wrote the senior essay all Yale political science majors admire. Moreover, David has been especially helpful in improving my research design.

Among the graduate students at Yale, I acknowledge the countless hours Yuke Li has put forth in discussions over dinner. I thank Yuke for being one of my closest friends during this research process. Also, I thank Mary McGrath for taking early version of my survey experiments and providing me with helpful comments. In addition, I think the undergraduates who have supported me. They include Vinicius Lindoso, Jack Newsham, Julia Fisher, Stacey Chen, Josh Kalla, Alyssa Bilinski, Josh Revesz, Ryan Pollock, the *Yale Daily News* Board of 2013, and the Political Science Senior Essay Colloquium.

I am particularly grateful for the support of my parents Hua Yang and Yueping Zhang. Unlike many Chinese parents who push their children into law, medicine, or business, they understand my desire to become a political scientist. I am truly grateful to have such open-minded and loving parents.

Finally, I dedicate my senior thesis to Peter Aronow, who is the best mentor a Yalie can wish for. When I walked into his section my junior year, I had no idea how much help he will be for my academic and personal life. Peter found me a job when I became unemployed and talked me out of a job when I was applying to graduate school. His amazing support and constant feedback kept me going through this senior thesis. Peter will become an amazing professor next year. I am lucky to have met him and I hope many future Yalies will have the same luck.

CONTENTS

I	INTRODUCTION	1
1	POLLING AT THE WATER'S EDGE	2
1.1	Introduction and plan of the senior thesis	2
1.2	The study of public opinion and foreign affairs	3
1.2.1	Public opinion and foreign policy	3
1.2.2	Experimental studies	5
1.3	Survey experiments: a methodological review	6
1.3.1	Overview of survey experiments	7
1.3.2	Pros and cons of survey experiments	9
1.4	The problem of confounding vignettes	11
1.4.1	Threat to construct validity	12
1.4.2	Directed acyclic graphs (DAGs)	13
1.4.3	Structural equation modeling framework	15
1.4.4	Placebo tests and two solutions	18
II	THREE EMPIRICAL STUDIES	20
2	DEMOCRATIC PEACE REVISITED	21
2.1	Introduction	22
2.2	Literature review	23
2.3	Hypotheses and methodology	25
2.3.1	Hypotheses	25
2.3.2	Survey experiment methodology	30
2.4	Results and discussion	33
2.4.1	Data	33
2.4.2	Placebo tests	35
2.4.3	Substantive findings	42
2.5	Conclusion	45
2.6	Acknowledgments	46
3	UN ENDORSEMENTS' EFFECT ON PUBLIC OPINION	47
3.1	Introduction	48
3.2	Literature review	49
3.3	Hypotheses and research design	51
3.3.1	Hypotheses	51
3.3.2	Research design	53
3.4	Results and Discussion	57
3.4.1	UN approval matters	58
3.4.2	Different intervention measures, different ATEs?	64
3.4.3	Who listens to the UN?	67
3.4.4	External validity check	72

3.5	Conclusion	73	
3.6	Acknowledgments	74	
4	FOREIGN OPPOSITION TO WAR AND THE AMERICAN PUBLIC		75
4.1	Introduction	76	
4.2	Literature review	78	
4.2.1	The influence of foreign elites	78	
4.2.2	Informative versus non-informative cheap talk		80
4.3	Hypotheses and research design	81	
4.3.1	Hypotheses	81	
4.3.2	Survey experiment design	83	
4.4	Results and discussion	86	
4.4.1	Data	87	
4.4.2	Effect of foreign opposition	88	
4.4.3	Heterogeneous treatment effects	93	
4.5	Conclusion	97	
4.6	Acknowledgments	98	
III	CONCLUSION	99	
5	INTERNATIONAL RELATIONS AND SURVEY EXPERIMENTS		100
5.1	Substantive contributions	100	
5.2	Methodological contributions	101	
5.3	Areas of future research	103	
IV	APPENDIX	105	
A	APPENDIX A: TEXTS OF THE SURVEYS		106
B	APPENDIX B: ADDITIONAL GRAPHS		125

LIST OF FIGURES

Figure 1	Treatment vignette that does not cause confounding	14
Figure 2	Treatment vignette that causes confounding	14
Figure 3	Vignette as valid versus invalid instruments	15
Figure 4	Log GNI by regime type	26
Figure 5	Log GNI per capita by regime type	27
Figure 6	Regime type map	28
Figure 7	Placebo test 1 by experiment type	37
Figure 8	ATEs by experiment type	44
Figure 9	Treatment tree	56
Figure 10	Support for intervention measures	60
Figure 11	Percentage of subjects supporting war by group	89
Figure 12	Percentage of subjects supporting war by group	90
Figure 13	Conditional average treatment effects	94
Figure 14	Naïve vignette type: distribution of support scores	126
Figure 15	Controls vignette type: distribution of support scores	126
Figure 16	Natural experiment vignette type: distribution of support scores	127

LIST OF TABLES

Table 1	Covariates balance check	36	
Table 2	Placebo tests: regions of the world	40	
Table 3	Placebo test: regions of the world	41	
Table 4	Estimating the ATE of democracy: without pre-treatment covariates	43	
Table 5	Estimating the ATE of democracy: with pre-treatment covariates		44
Table 6	Treatment/control chart	55	
Table 7	Covariates balance check	59	
Table 8	Estimating the treatment effects of UN endorsements		61
Table 9	GLS Regression: Random Effects	63	
Table 10	GLS regression: random effects; by intervention types		65
Table 11	Three groups with similar baseline support	66	
Table 12	GLS regression: random effects; three groups with similar baseline support	66	
Table 13	Conditional average treatment effects (CATEs)	68	
Table 14	GLS regression: random effects; heterogeneous treatment effects		69
Table 15	Covariates balance check	88	
Table 16	Effect of treatments on Support Scores	91	
Table 17	Effect of treatments on support scores	92	
Table 18	Democrats: effect of treatments on support scores		95
Table 19	Republicans: effect of treatments on support scores		96

Part I

INTRODUCTION

1 POLLING AT THE WATER'S EDGE

1.1 INTRODUCTION AND PLAN OF THE SENIOR THESIS

This senior thesis stands at the intersection of international relations (IR) and American politics. Are Americans less likely to go to war against democracies than autocracies? How do international organizations affect public opinion about military interventions and economic sanctions? And how do Americans react to foreign opposition to war? These are the substantive questions my work sets out to explore. But at its core, this thesis is also about ways to make accurate causal inference from survey experiments. Pollsters and scholars alike have used surveys to understand the questions posed above; however, until the recent introduction of experimental methods, researchers had a difficult time separating correlation from causation due to endogeneity. Randomly assigned treatments embedded in surveys seem to solve the problem of endogeneity. Nevertheless, as I will demonstrate in this thesis, many treatment vignettes¹ in survey experiments cause confounding by affecting subjects' beliefs along not one but multiple dimensions.

In order to estimate unbiased treatment effects in survey experiments, scholars should evaluate the quality of the vignettes they assign. Furthermore, they need to develop strategies to insure that 1) the assignment of treatment vignettes is random *and* 2) the experimental element in the treatment vignettes appear as-if random. To meet the latter criterion, I propose two strategies. First, one can include additional details in experimental vignettes to control for confounding factors. Second, one can invent hypothetical natural experiments and embed them in the vignettes.

Part 1 of my thesis discusses the evolution of research on public opinion and foreign affairs. The first chapter begins with a review of ways theorists and empiricists from international relations and American politics have conceptualized voters' attitudes towards

¹ I define a treatment vignette as the wording in a survey experiment prompt or question.

international affairs. Next, I discuss the advantages and disadvantages of using survey experiments for public opinion research. After presenting existing critiques, I formalize the problem of confounding vignettes in survey experiments and provide two solutions to fix this problem.

Part 2 of my thesis (Chapter 2-4) consists of three empirical studies that showcase some of my methodological innovations. Findings from these three studies also provide substantive knowledge about the conditions under which Americans support war. The second chapter replicates and expands previous survey experiments testing the democratic peace in a public opinion context. The third chapter examines whether UN Security Council endorsements affects public support for military interventions and economics sanctions. The fourth chapter studies whether Americans heed foreign elites' opposition to war.

Finally, in the Conclusion (Part 3), I summarize key results and explicate the substantive and methodological implications of my work. In addition, I discuss how my findings relate to new areas of research.

1.2 THE STUDY OF PUBLIC OPINION AND FOREIGN AFFAIRS

1.2.1 *Public opinion and foreign policy*

The study of U.S. public opinion and foreign policy is filled with debates. Two major questions arise: Do Americans have coherent attitudes about foreign policy? And, even if Americans have opinions about foreign policy, what does it matter for national leaders?

One view of American public opinion, tracing its roots to the "Michigan School," holds that voters, in general, have incoherent and unsophisticated political attitudes. In *The American Voter*, Campbell et al. report that only a small portion of the electorate consider policies and issues when voting. Instead most Americans vote based on their partisan affiliations (Campbell et al. 1960). Likewise, Converse finds that most voters do not understand political issues and do not behave like ideologues (Converse 2006). Uninformed voters taking

cues from party elites remains conventional wisdom in the study of public opinion and public policy (Zaller 1992, Lenz 2012). Foreign policy is no exception: Berinsky (2007) uses observational data and survey experiments to demonstrate that elite discourse affect the mass public's evaluation of wars' success or failure.

Nevertheless, some scholars argue that American voters hold their own foreign policy preferences and evaluate leaders based on those preferences. Survey experiments have demonstrated that voters evaluate whether the U.S. should go to war based on their dispositions (such as isolationism versus internationalism), geopolitical context of the conflict, and the interaction between these two factors (Herrmann, Tetlock & Visser 1999). Evaluation of the American National Election Studies (ANES) data in the late 1980s suggests that Americans considered national security a salient issue and evaluated candidates' positions on national security (Aldrich, Sullivan & Borgida 1989). Nincic and Hinckley's (1991) study suggests that voters evaluate presidential candidates' positions on foreign policy as much their positions on domestic policy. In more recent times, scholars have studied the effects of Iraq War fatalities on national elections. Looking at state-level data, Karol and Miguel (2007) show that Iraq casualties significantly decreased Bush's vote share and predicted that Bush might have won two percent more of the popular vote if there had been no casualties. Likewise, Gelpi, Reifler, and Feaver's (2009) analysis of survey data suggests that voters' evaluation of the Iraq War's likelihood of success and their beliefs about the war's justifications are significant predictors of vote choice. These studies argue public opinion about salient foreign policy, such as major wars, matter because of the electoral connection.

In my senior thesis, I do not assert whether the public is rational or irrational, sophisticated or unsophisticated. Instead, through survey experiments, I study how the mass public are affected by their prior beliefs as well as the globalized society around them. Unlike previous studies that examine how domestic elites influence public opinion, I study how outside actors, such as the United Nations (UN) and foreign countries, affect Americans' attitudes towards war. If we believe the American voters can hold their leaders accountable

through elections, then public opinion about wars does matter — not only for the U.S. but also for the world at large.

1.2.2 *Experimental studies*

As scholars of American politics and comparative politics are increasingly using experimental methods, IR researchers, too, appear to be moving in this direction. While field experiments have become common within political science, IR scholars are constrained in applying real-world interventions. Instead, they have largely relied on survey experiments and lab experiments (Mintz, Yang & McDermott 2011). According to Roth's (1995) review, experimental works serve three main purposes: 1) testing theoretical models, 2) generating data, and 3) evaluating potential policies. Since IR has largely been dominated by theories, most experimentalists in the discipline focus on testing formal and informal theoretical models.

The first generation of IR experiments consists of cooperation games or simulated international conflicts in lab settings (Deutsch et al. 1967, Majeski & Fricks 1995, Pilisuk 1984, Beer et al. 1995). In contrast, the new generation of IR experiments has shifted the focus away from elite decision-making towards public opinion. They seek to test theories that involve mass perception of war, such as domestic audience cost (Tomz 2007) or the democratic peace (Mintz & Geva 1993). In these types of survey experiments, respondents are presented with different versions of a vignette, depending on whether they are in the control or treatment group. After reading the vignettes, respondents answer one or more questions that measure their attitudes. Tomz's (2007) study of domestic audience cost utilized 1,127 adults, representative of the U.S. population, to test whether voters punish the president if he were to back down after making foreign threats. Through his survey experiments, he uncovers a substantial drop (16 percentage points) in public approval if the president does not follow through with his threat. Mintz and Geva's (1993) evaluation of democratic peace involved 117 subjects from three samples of American and Israeli col-

lege students and U.S. adults. Respondents were presented with vignettes in which one hypothetical country has invaded another hypothetical country. Half of the subjects read that the invader was a stable democracy and the other half read the invader was a military dictatorship. Subjects were more likely to favor use of force against the invader when it was a military dictatorship than when it was a democracy.

Other scholars have moved away from testing well-established IR theories to evaluating notable trends in public opinion and foreign policy. For instance, Grieco et al.'s (2011) study examined whether endorsements by international organizations affect public support for war. Through their survey experiment, they discover that UN approval substantially made subjects more willing to support humanitarian interventions.

Moving beyond the traditional discussions of war and peace, some scholars have begun to seriously consider the intersection of foreign policy and domestic politics. For example, Hainmueller and Hiscox's (2010) survey experiment of opposition to immigration evaluates whether economic concerns generate anti-immigrant sentiments among American citizens. They discover American – rich and poor alike – overwhelmingly favor highly skilled immigrants over low-skilled immigrants. In states with high fiscal exposure, poor natives, compared to rich natives, are more opposed to low-skilled immigration. These conflicting results suggest that economic self-interest does not explain a large part of anti-immigration attitudes.

Building upon these existing IR survey experiments, I seek to test political theory and findings from observational studies.

1.3 SURVEY EXPERIMENTS: A METHODOLOGICAL REVIEW

Beyond questions substantive questions about American voters' reactions to foreign affairs in my senior thesis, I also consider the methodological strengths and shortcomings. In this section, I provide an overview of the different types of survey experiments. Next, I addresses the pros and cons of this method, including existing criticisms.

1.3.1 Overview of survey experiments

Although survey experiments have existed since the 1950s, it only became a popular methodology among social scientists in the late 1990s with the rise of computer-assisted telephone interviewing. Proponents of survey experiments often claim that this strategy overcomes the problems of traditional, non-experimental surveys. Instead of running regressions with dozens of controls to minimize confounding, researchers purportedly can now unravel cause and effect by randomly assigning treatments and observing outcomes (Brady 2000, Gilens 2002, Mutz 2011).

Gilens (2002) reviews the different types of survey experiments. The simplest and most common type involves changing question wording. One prominent example is the “welfare mother experiment” on the 1991 Race and Politics Survey conducted by the University of California, Berkeley.² Survey subjects were randomly sorted into two experimental groups. Each subject was given a vignette about a hypothetical welfare mother; keeping other features of the mother constant, the researchers told the control group that she is white and the treatment group that she is black. Then subjects were asked to predict the behavior of the welfare mother and other questions about race and welfare. Other examples include surveys that manipulate the race of criminals (Hurwitz & Peffley 1997), the gender of political candidates (Sanbonmatsu 2002), the branch of government that made a policy decision (Gibson, Caldeira & Spence 2005), the race of immigrants (Schildkraut 2009), and the party that endorses a particular foreign policy (Trager & Vavreck 2011).

The second type of survey experiments, the “framing” or “priming” experiments, involves changing the question context. An often-cited example is Sniderman and Carmine (1997) “mere mention” study, reported in *Reaching Beyond Race*. Subjects were randomly assigned to two experimental groups. The ones in the control group were asked to rate blacks as a group on a series of traits and stereotypes. The ones in the treatment group were first asked a question about affirmative action and then proceeded to the traits and

² The 1991 Race and Politics Survey was directed by Paul M. Sniderman, Philip E. Tetlock, and Thomas Piazza with support from the National Science Foundation (SES-8508937).

stereotype questions. Sniderman and Carmine interpret the significant difference between the two groups' responses to the stereotype battery means that merely mentioning affirmative action can "prime" negative perceptions of blacks. Other examples include priming subjects with correct news information (Gilens 2001), political arguments about immigration (Sniderman, Hagendoorn & Prior 2004), and messages from political elites (Druckman & Nelson 2003).

The third type of survey experiments, the list experiment, is used to detect subjects' true sentiments on sensitive topics (Kuklinski 1997). For instance, subjects might feel uncomfortable disclosing choices that might appear racist in a survey. In the list experiment, the respondents hear a list of items that they are told "might make people angry or upset." The respondents were told to indicate how many of the items — not which ones — make them angry or upset. In the control version, the subject is given four items. In the treatment version, the subject is given five items; the additional item is the experimental treatment of interest. By comparing the mean number of upsetting items in the baseline condition (4 items) with the treatment condition (5 items), researchers can calculate the percentage of the sample population who express hostility towards the item of interest.

For the purposes of my senior thesis, I will examine confounded vignettes in the first type of survey experiments, the ones in which researchers change question wording. I do so for two reasons. First, this is the most common type of survey experiments. Second, the treatment and control conditions in these experiments are usually identical except one or a few words. Therefore, intuitively, one would think these vignettes would cause minimum confounding. Nevertheless, I argue otherwise and am willing to accept a high burden of proof.³

³ It is plausible that "framing" or "priming" experiments could also experiment confounding. The treatments researchers administers could prime subjects in different ways than the experimenters intend.

1.3.2 *Pros and cons of survey experiments*

Proponents of survey experiments advocate for the methodology mainly because they think it overcomes the problem of endogeneity found in data from non-experimental surveys. “Because respondents are randomly assigned to treatment groups, we know that — within the bounds of sampling error — the subgroups are identical in every way...Thus the problem of associations and ambiguous causal relations that plague cross-sectional survey analysis are avoid,” Gilens writes (2002, 248). Furthermore, he argues that researchers might present their theories *ex ante*, minimizing the possibility of data mining common to those who use big omnibus surveys. Likewise, Brady (2000, 52) bemoans the utility of non-experimental surveys: “[they] have not been much use to researchers studying the impact of political events and contexts...[survey experiments], however, made it possible for political scientists to obtain the increased control need for testing theories.” Furthermore, Brady writes survey experiments fielded to a nationally representative sample have more external validity than lab experiments, which often use college students as subjects. Similarly, Mutz (Mutz 2011, 15) argues that “in addition to providing a means of resolving direction of causation and potentially spurious relationships, population-based experiments can help advance theory in research areas where selection bias makes observational studies relatively unproductive.” In sum, these advocates of survey experiments suggest that this method avoids the problem of confounding because treatments are randomly assigned.

Gaines, Kuklinski and Quirk (2007) provide a comprehensive overview of existing critiques of survey experiments. The five main criticisms the authors summarize are:

1. short-lasting treatment effects
2. looking mutual causation
3. lack of control groups
4. spillover within a survey
5. spillover from the real world

The main critique of survey experiments centers on their lack of external validity. As the authors report, several studies—including Luskin, Fishkin, and Jowell (2002) and Druckman and Nelson (2003)—have found that survey experiment treatments are short-lived. Furthermore, Barabas and Jent’s (2010) study of public opinion of Medicare and immigration shows that information provided in survey experiments increased political knowledge and changed attitudes. But real-world announcements had no discernible effects; even those exposed to facts by the mass media showed smaller effects than the treated subjects in the survey experiments.

Gaines, Kuklinski and Quirk’s concerns about spillover effects in survey experiments are of particular interest to me. The authors recognize that treatments in survey experiments can be influenced by the way the subjects are exposed to facts in the real world. For instance, suppose researchers are unaware that all their subjects were exposed to political messaging about affirmative action through some major news event prior to the experiment. The researchers then conduct a survey experiment to test how framing about affirmative action can affect political attitudes. Because subjects assigned to control received real world exposure to the affirmative action framing, the average treatment effect (ATE) of the survey experiment treatment would be artificially depressed.

One aspect of the authors’ critiques address the problem of confounded treatment vignettes in survey experiments. Gaines, Kuklinski and Quirks write that researchers often do not have defined controls. As an example, they discuss a hypothetical survey experiment that asks whether jobs should be set aside for one of two minority groups. They write, “if respondents are randomly asked about jobs set aside either for African-Americans or for Mexican Americans, some respondents may defy the contrast by inferring that such programs generally cover both blacks and Mexican Americans, despite no explicit mention of the other group” (Gaines, Kuklinski & Quirk 2007, 17). In the naïve design, researchers assume the vignette only changes subjects’ perception about affirmative action quotas for African-Americans. In reality, the vignette might also change subjects’ perceptions about affirmative action quotas for other underrepresented minorities. This happens because sub-

jects have some exposure to facts in the real world; in the real world, affirmative actions programs usually help multiple underrepresented minority groups. Therefore, it would be difficult to interpret the difference in attitudes between the two groups as the difference between subjects' preference for one ethnicity over another.

1.4 THE PROBLEM OF CONFOUNDING VIGNETTES

A big proportion of my senior thesis explores the problem of confounding vignettes in survey experiments. Before proceeding to an empirical study of this problem, I define it in three ways. First, I discuss this problem as a threat to construct validity. Second, I demonstrate the problem using directed acyclic graphs (DAGs). Third, I present this problem as a violation of the exclusion restriction in a structural equation modeling framework. Finally, I explain how to use placebo tests to detect the problem of confounding vignettes and present two design-based solutions to overcome this problem. The first solution involve adding controls to the treatment vignettes so they are more precise. The second solution is to invent a natural experiment and insert it into the treatment vignettes to make the experimental element in them appear as-if random.

Before I dive into this methodologically problem, I should note that researchers randomly assign vignettes when conducting survey experiments. Therefore, I do not mean that vignette assignment is confounded. Instead, I mean that the wording of treatment vignettes is often not precise and could change subjects' beliefs in ways experimenters do not intend to. For instance, in a survey experiment about the democratic peace, subjects are told that the aggressor country is either a "democracy" or a "dictatorship." Because the treatments only mention the aggressor country's regime type, subjects can interpret the other characteristics of the country. If subjects know facts about the real world, they would associate democracies with countries that have developed economies and an European location. Likewise, they would associate dictatorships with countries that have un-

derdeveloped economies and Subsaharan African location. Therefore, the naïve vignette of “democracy” versus “dictatorship” could cause confounding in subjects’ mind.

1.4.1 *Threat to construct validity*

One way to conceptualize the problem of confounding vignettes in survey experiments is to consider them as a threat to construct validity. A construct is a hypothesis that researchers attempt to test through operations in an experiment (Campbell & Stanley 1963, Shadish & Cook 2002). In our democratic peace example, the construct we want to study is whether Americans are more or less likely to support going to war against a democratic country than an autocratic country. We operationalize this construct in our survey experiment by varying the regime type (e.g., “democracy” or “dictatorship”) in the vignettes. Construct validity is achieved when the operation in the experiment closely represents the constructs one wishes to test.

The particular threat to construct validity in the democratic peace example is construct confounding. As Shadish and Cook writes, “Operations usually involve more than one construct, and failure to describe all the constructs may result in incomplete construct inferences” (2002, 73). Although we merely vary one word in our operation, “democracy” and “dictatorship” involve more than just the construct of regime type. Other constructs the operation could potentially include are the aggressor country’s level of economic development and geographic location.

Suppose we conduct our survey experiment and discover that subjects given the “democracy” vignette were significantly less likely to go to war against the aggressor country than those given the “dictatorship” vignette. From this result, can we draw the inference that Americans vary their support for war against countries based on regime types? Not necessarily. Unless we can be sure that our operation closely represent our construct, we cannot draw such an inference. Some subjects may have lowered their support for war against the “democracy” because they think democracies have high trade with the U.S. or are more

culturally similar to the U.S. Therefore, before one can draw unbiased inference from survey experiments, one should make sure that the operation in one's survey experiment does not valid construct validity.

1.4.2 Directed acyclic graphs (DAGs)

Directed acyclic graphs (DAGs) are commonly used in epidemiology to illustrate causal pathways between variables. I use DAGs to illustrate how survey experiment vignettes affect subjects' beliefs and an outcome measure. Researchers common consider the model of causality outlined in Figure 1: a randomly assigned treatment vignette Z changes subjects' beliefs along the dimension of interest T and induces changes in the outcome measure Y . Experiments keep the control and treatment vignettes the same except for one element because they want to change subjects' beliefs only along the dimension of interest. For instance, in a survey experiment about democratic peace, experiments only vary the regime type of the aggressor country and not other characteristics. If the model in Figure 1 holds, the ATE can be calculated as $E[Y_i(\text{Vignette}_{\text{treatment}}) - Y_i(\text{Vignette}_{\text{control}})]$.

Nevertheless, survey experiment vignettes could potentially change subjects' beliefs about other factors, which in turn, affect the outcome Y as illustrated in Figure 2. In Figure 2, the vignette Z not only changes subjects' belief along the dimension of interest T_1 but also along other dimensions $T_{2,\dots,n}$. Confounding would not be a problem if $T_{2,\dots,n}$ has no effect on the outcome Y , but we cannot easily make that assumption about every $T_{n>1}$. For instance, In the democratic peace survey experiment, if the naïve vignette also affects subjects' beliefs about the wealth of the aggressor country, subjects will most likely incorporate the wealth of the aggressor country into their calculus when deciding how to respond to the outcome measure question.

One way to think about vignettes in survey experiments is to consider them as instruments for some treatment T . In this model, as depicted in Figure 3, the vignettes (i.e., words in a survey question) are not the actual treatment. Instead the treatment T is the

Figure 1: Treatment vignette that does not cause confounding

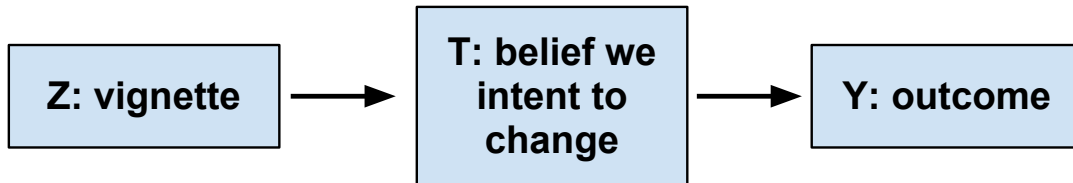
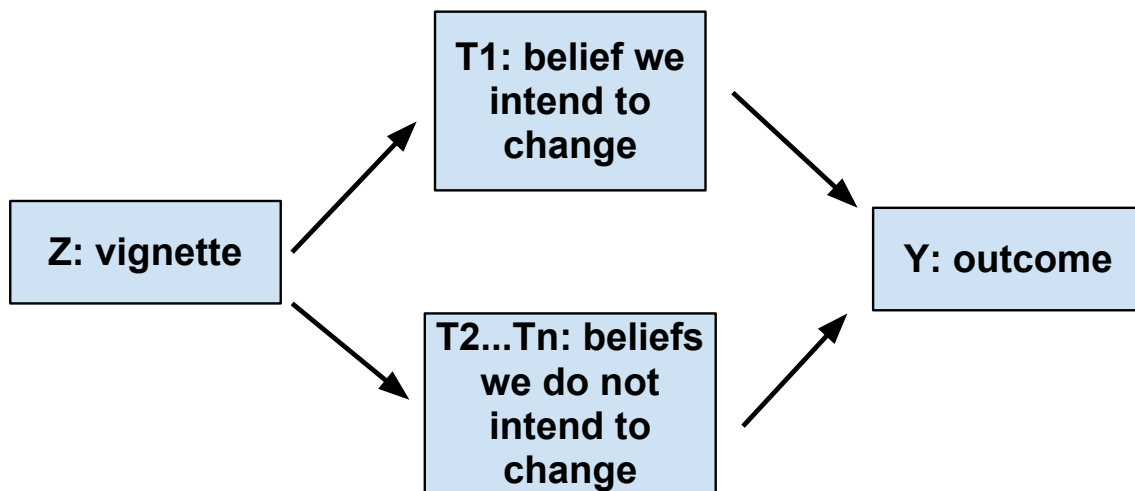


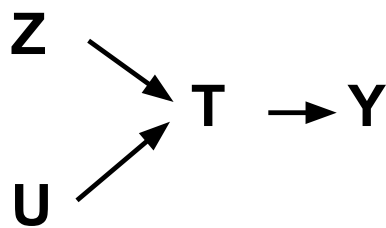
Figure 2: Treatment vignette that causes confounding



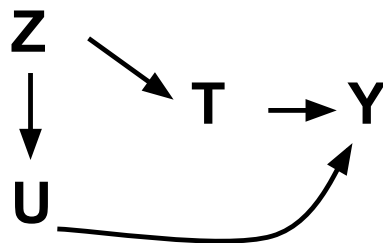
belief of subjects the researcher wants to change. The vignette Z is a valid instrument for T if and only if Z only affects Y through T and not through some back channel U . In the next subsection, I advance this model of causality within a structural equation modeling framework.

Figure 3: Vignette as valid versus invalid instruments

vignette Z as a valid instrument for treatment T



vignette Z as an invalid instrument for treatment T



1.4.3 *Structural equation modeling framework*

Now, I explain the problem of confounding vignettes within a structural equation modeling framework. Generally, political scientists do not conceptualize survey experiments in terms of instrumental variables. They usually conceive a simple model of a treatment T affecting some outcome Y : $T \rightarrow Y$.

Instead, I consider the treatment vignette Z as an instrument affecting subjects' belief along one particular dimension of interest. We should consider the change in belief along the intended dimension as the actual treatment T . T , in turn, affects the outcome variable Y . Our model is such: $Z \rightarrow T \rightarrow Y$. A vignette Z is a valid instrument if and only if it induces a change in Y through T . Otherwise, the treatment vignette Z would violate the exclusion restriction and cause confounding.⁴

Suppose there exists a linear relationship between T_i , subjects' belief along the dimension of interest to experimenters, and Y_i , the outcome variable. In our democratic peace example, T_i is each subject's perception of Country A 's regime type (0 if autocracy and 1 if democracy) and Y_i is each subject's support for war against Country A . The relationship between T_i and Y_i is such:

$$Y_i = \beta_1 + \beta_2 T_i + u_i \tag{1}$$

Suppose that for some reason T_i has a random component that is not distributed independently of u_i . In the democratic peace example, unobserved factors in u_i could include each subject's perception of Country A 's level of economic development or geographic location. In the real world, T_i and u_i might not be independently distributed since people think democratic countries tend to be economically developed and/or located in the Global North. Therefore, an OLS regression of Y on T would lead to a biased estimate of T 's effect on Y .

A way to get around this problem using a survey experiment is to construct a pair of vignettes about an aggressor Country A that only varies Country A 's regime type. Subjects can then be randomly assigned to read either the democracy vignette or the autocracy vignette. Furthermore, we can think about the vignette as an instrument Z for T , subjects' beliefs about Country A 's regime type. The instrument Z should be correlated with T but not correlated with u according to the exclusion restriction. One can safely assume

⁴ I arrived at this model independently from other scholars, although John Bullock has also been working within this framework.

that correlation between Z and T is non-zero⁵. But one cannot readily assume that Z is uncorrelated with u because the vignette can cause confounding. Recall that u include each subject's perception about characteristics of Country A aside from its regime type. The labels "democracy" or "autocracy" in the vignette Z introduce other connotations about Country A that might affect subjects' support for war against Country A.

Suppose one naively thinks that Z is uncorrelated with u , so one uses Z as an instrumental variable to estimate the effect of T on Y . The IV estimator, denoted as b_2^{IV} , is:

$$b_2^{IV} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^n (Z_i - \bar{Z})(T_i - \bar{T})} \quad (2)$$

Substituting from Y from Equation 1, I expand the expression for b_2^{IV} :

$$\begin{aligned} b_2^{IV} &= \frac{\sum_{i=1}^n (Z_i - \bar{Z})[(\beta_1 + \beta_2 T_i + u_i) - (\beta_1 + \beta_2 \bar{T} + \bar{u})]}{\sum_{i=1}^n (Z_i - \bar{Z})(T_i - \bar{T})} \\ &= \frac{\sum_{i=1}^n [\beta_2 (Z_i - \bar{Z})(T_i - \bar{T}) + (Z_i - \bar{Z})(u_i - \bar{u})]}{\sum_{i=1}^n (Z_i - \bar{Z})(T_i - \bar{T})} \\ &= \beta_2 + \frac{\sum_{i=1}^n (Z_i - \bar{Z})(u_i - \bar{u})}{\sum_{i=1}^n (Z_i - \bar{Z})(T_i - \bar{T})} \end{aligned} \quad (3)$$

Note that the IV estimator is equal to the true effect of T on Y (β_2) and an error term. To see the effect of sample size n on the IV estimator, I take the probability limit of b_2^{IV} :

$$\text{plim } b_2^{IV} = \beta_2 + \frac{\text{plim } \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(u_i - \bar{u})}{\text{plim } \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(T_i - \bar{T})} \quad (4)$$

$$= \beta_2 + \frac{\sigma_{Zu}}{\sigma_{ZT}} \quad (5)$$

If Z is distributed independently of u , then $\sigma_{Zu} = 0$. Therefore, in large samples, b_2^{IV} would produce an unbiased estimate of T 's effect on Y . But if Z is not distributed inde-

⁵ That is one can safely assume that telling subjects Country A is a democracy will mostly likely make them perceive Country A as a democracy. Although if one is unsure, one can check that Z and T are correlated by performing a manipulation check.

pendently of u , in the case of confounding vignettes, then $\sigma_{X_u} \neq 0$. Therefore, no matter what the sample size, b_2^{IV} would produce a biased estimate of T 's effect on Y . Note that the treatment effect I estimate is the local average treatment effect (LATE), or the average treatment effect among those who comply with treatment assignment. This is not a problem in online survey experiments since the researcher can simply restrict access to each version of the survey based on treatment assignment.

1.4.4 *Placebo tests and two solutions*

To detect the problem of confounding in survey experiments, I perform placebo tests after subjects read the vignettes. The purpose of the placebo tests is to measure whether the vignettes have affected subjects' beliefs beyond the factor of interest. In the democratic peace example, for instance, I ask subjects whether it is plausible that the conflict between the aggressor country and its neighbor took place in the developing world. Perceptions of where the conflict took place should not be affected by the differing vignettes of regime type. However, if they are affected, then we must revise our vignettes to minimize confounding.

I propose two design-based strategies to minimize confounding in vignettes: 1) holding confounds fixed, analogous to controlling for confounds by including additional details in the vignettes, and 2) exploiting variation in the vignettes that is, perceived by the respondents to be, as-if random. The first "controls" strategy will succeed to the extent that confounds can be fully specified in the vignettes. Like studies of observational data, adding controls is often not sufficient. In the second strategy, I embed a plausible natural experiment in the scenario to make the difference between the treatment and control vignettes appear as-if random. If the respondents perceive the variation in vignettes to be as-if random, then their beliefs about other features should be independent of vignette assignment. I propose using placebo tests to evaluate whether one's scenario succeeds in overcoming confounding.

I examine the problem of confounding in survey experiments by replicating and expanding upon a set of existing survey experiments (Mintz & Geva 1993, Tomz & Weeks 2012a). In previous survey experiments about the democratic peace, simply-worded vignettes are likely to cause confounding. For instance, respondents might associate democracy with higher levels of economic development or some geographic regions. I implement the two strategies for overcoming confounding, evaluate the extent to which they succeed through placebo tests, and reassess the results in light of our findings.

In addition, I conduct two additional survey experiments where the vignettes do not cause confounding. In these experiments, the vignettes fully operationalize the constructs I seek to test and do not pose threats to construct validity. First, I test whether United Nations Security Council resolutions influence American public opinion about going to war. Second, I examine whether Americans respond to foreign elites who oppose a U.S.-backed war. Results from these survey experiments suggest that a proportion of Americans are indeed responsive to foreign elite cues.

Part II

THREE EMPIRICAL STUDIES

2 DEMOCRATIC PEACE REVISITED

Abstract. Scholars of international relations have noted that democracies tend to engage in conflicts less often with other democracies than with autocracies. Several survey experiments have confirmed this observation. Nevertheless, a naïve experimental vignette that merely changes the regime type of the aggressor country might cause confounding by affecting subjects' beliefs about the aggressor country apart from its regime type. Using three sets of survey experiments, I compare the naïve design with vignettes that add controls and vignettes with a "natural experiment" embed. While the more complex vignettes performed better in placebo tests, the naïve design caused less confounding than predicted. But contrary to results from previous survey experiments that confirm the democracy peace theory, I detect significant effect of democracy in only one of the three vignette types.

2.1 INTRODUCTION

The theory of democratic peace, aside from its philosophical importance, is one of the most empirically robust findings in international relations (IR). One approach is to find associations between democratic regime type and peace among country-dyads through statistical analysis of observational data. Another approach uses survey experiments to test whether subjects in democracies profess greater support for war against autocracies versus democracies. Proponents of survey experiments argue that one can draw unbiased causal inference because the treatments are exogenous, that is they are uncorrelated with unobserved factors that also affect the outcome variable. But treatment vignettes in survey experiments about democratic peace are simply constructed and could affect subjects' beliefs about the aggressor country apart from its regime type. The wording of differing regime type (democracy vs. autocracy) is likely confounded with unspecified features of democratic states and autocratic states. For instance, respondents might associate autocracies with being economically underdeveloped or located in Sub-Saharan Africa.

The democratic peace project consists of three mini-survey experiments.¹ In the first experiment, I use a naïve vignette in which I only vary the regime type of the aggressor country. I then test for confounding through placebo tests. In the second experiment, I hold confounds fixed by telling subjects details about the aggressor country (e.g., whether the country is developed or underdeveloped). In the third experiment, I insert invented “natural experiments” in the vignettes to make the difference between regime types as-if random. I hypothesize that the naïve vignette causes the most confounding, the vignette with controls causes the second most confounding, and the natural experiment embed vignette causes the least confounding.

The results from a pilot study are somewhat inconsistent with my hypotheses. They suggest the naïve vignette is not as confounded as I imagine. The naïve treatment vignette passed the placebo test question that detects economic development as a confounder.²

¹ This project is a collaboration with Allan Dafoe, who originally provided me with the idea of confounding vignettes in survey experiments, and Devin Caughey.

² A treatment vignette passes a placebo test when subjects given the treatment scenario answer the placebo test question the same way as subjects given the control scenario.

Nevertheless, the naïve vignette failed to past some placebo tests designed to detect geographic location as a confounder. The substantive findings from my survey experiments are very different compared with results from previous survey experiments on democratic peace. In the experiments with the naïve vignette and the vignette with controls, subjects were equally likely to support war against democracies and dictatorships. Only subjects in the experiment with the natural experiment vignette expressed significantly less support for going to war against democracies than autocracies.

2.2 LITERATURE REVIEW

Since Kant put forth a nascent idea of the democratic peace in his (1795) essay *Perpetual Peace*, IR scholars have debated whether democracies are less likely to fight democracies than autocracies. On the proponent side, scholars have used regression analyses to show that there exists a significant association between dyadic democracy and peace (Dafoe 2011, Gartzke 1998, Kacowicz 1995, Lemke & Reed 1996, Maoz & Abdolali 1989, Maoz & Russett 1993, Oneal & Russett 1999b, Ray 1995, Rousseau et al. 1996, Russett 1993, Russett & Oneal 2001, Russett, Oneal & Davis 1998, Signorino & Ritter 1999, Melvin & Singer 1976, Thompson & Tucker 1997). But some critics offer alternative explanations for the democratic peace. One prominent alternative is the “capitalist peace,” which suggests that free markets and economic development – not democratic regimes – account for peace between countries (Gartzke 2007, Mousseau 2000, Mousseau, Hegre & Oneal 2003). Aside from performing statistical analyses on macrohistorical data, another way to empirically test the democratic peace theory is to conduct survey experiments in which subjects are randomly assigned vignettes about different regime types.

Experimental approaches to the democratic peace is not new: there exists five survey experiments on this subject. Mintz and Geva’s (1993) used American college students, average American voters, and Israeli students, with a total of 117 subjects. The two regime types tested were a stable democracy and a “military dictatorship.” Subjects were

more likely to favor using force when the aggressor country was a military dictatorship. Rousseau (2005) conducted similar experiment on 141 American college students. The two regime types Rousseau used were a “democratically elected government” and a “single-party dictatorship.” Likewise, support for using military force against the the democracy is much lower than against the dictatorship.

Johns and Davies’s (2012) study used large N samples representative of the U.S. and UK populations (2000+ subjects in each country). They not only tested whether subjects would respond differently to democracies versus autocracies but also to the majority religion of the aggressor country (Christian versus Muslim). Their results once again confirms the democratic peace theory; in addition, they found that subjects were more likely to support attacking the Muslim country than the Christian country.

Finally, Tomz and Weeks (2012a, 2012b) have recently completed a set of survey experiments (large N, nationally representative sample of Americans and Britons) that look into the mechanism of democratic peace. In the first set of survey experiments, the two regime types were a) a democracy that is unlikely to become an autocracy and b) an autocracy that is unlikely to become a democracy. In addition, Tomz and Weeks controlled for the aggressor country’s military capabilities and alliances. They discover that subjects are more likely to support military strikes against autocracies than democracies. Tomz and Weeks’s second set of experiments test the substance behind the democratic peace. The experimental conditions they create include a) whether democratic elections were held in the aggressor country and b) whether the aggressor country violates human rights. They found that human rights practices are even more powerful than democratic institutions (i.e., free and fair elections) in explaining subjects’ decision to use force or not.

Except for Tomz and Weeks’s studies, none of the authors above look into whether their vignettes are confounded with other unobserved factors. The regime type treatments they administer are rather simple and draw explicit lines between the democracy and the autocracy. There are no as-if random assignment of regime types: the democracies have been and will be stable for a long time and the autocracies likewise.

2.3 HYPOTHESES AND METHODOLOGY

In this section, I first outline the main hypotheses that I test in my survey experiments. I then explain the three sets of mini-survey experiments I conduct to detect confounding in treatment vignettes.

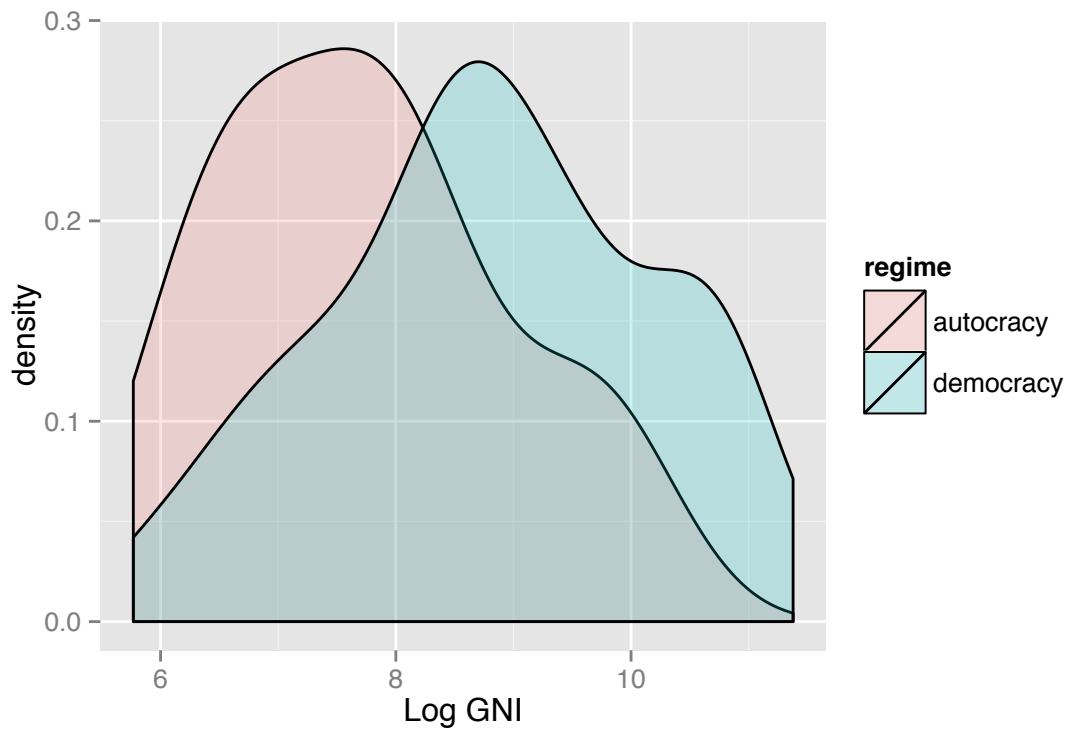
2.3.1 *Hypotheses*

There exist several reasons why naïve treatment vignettes in survey experiments testing democratic peace could cause confounding. The change of a single world, from “democracy” to “dictatorship,” not only changes subjects’ beliefs about the regime type of the country but also could change their beliefs about other characteristics of the country. In the real world, democracies and autocracies are dissimilar in many ways. I point out two such salient differences that could affect subjects’ perceptions.

First, democracies tend to be richer than autocracies. Figure 4 and 5 show the distribution of wealth (both GNI and GNI per capita) by democracies and autocracies. The Quality of Government dataset include a 10-point score of countries by levels of democracy, with 10 being the most democratic and 0 being the least democratic. I sort countries using the Quality of Government score: countries with score of 7 and above are labeled democracies and countries with score of 3 and below are labeled autocracies. As the two figures show, while there exists some overlap between democracies and autocracies, the distribution is bi-modal. Democracies, on average, tend to have greater log GNI and log GNI per capita. If subjects know something about the distribution of wealth by regime type, then they might use this information when responding to the naïve vignettes.

Second, democracies and autocracies are not uniformly distributed in terms of locations around the world. As Figure 6 demonstrates, certain regions of the world have higher concentrations of autocracies than others. For instance, central Asia, the Middle East, and Sub-Saharan Africa contain many autocratic countries. In contrast, western Europe, North

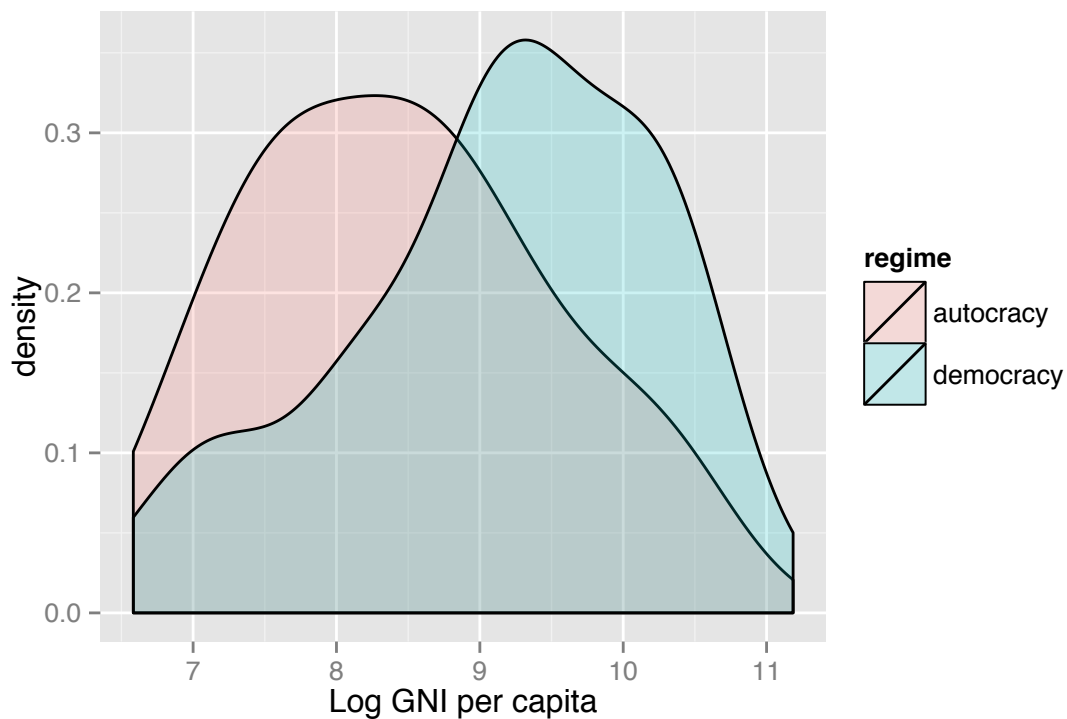
Figure 4: Log GNI by regime type



Data from Quality of Government dataset

America, South America and Oceania mostly consist of democracies. While subjects may not have an exact map like Figure 6 in their minds, they have some knowledge about the distribution of autocracies and democracies in the world. Therefore, when they see the word “democracy” or “dictatorship” in the naïve vignette, they would associate the aggressor country with a certain geographic region. If subjects like some regions better than others (due to ethnocentrism or racism), the naïve vignette would cause confounding.

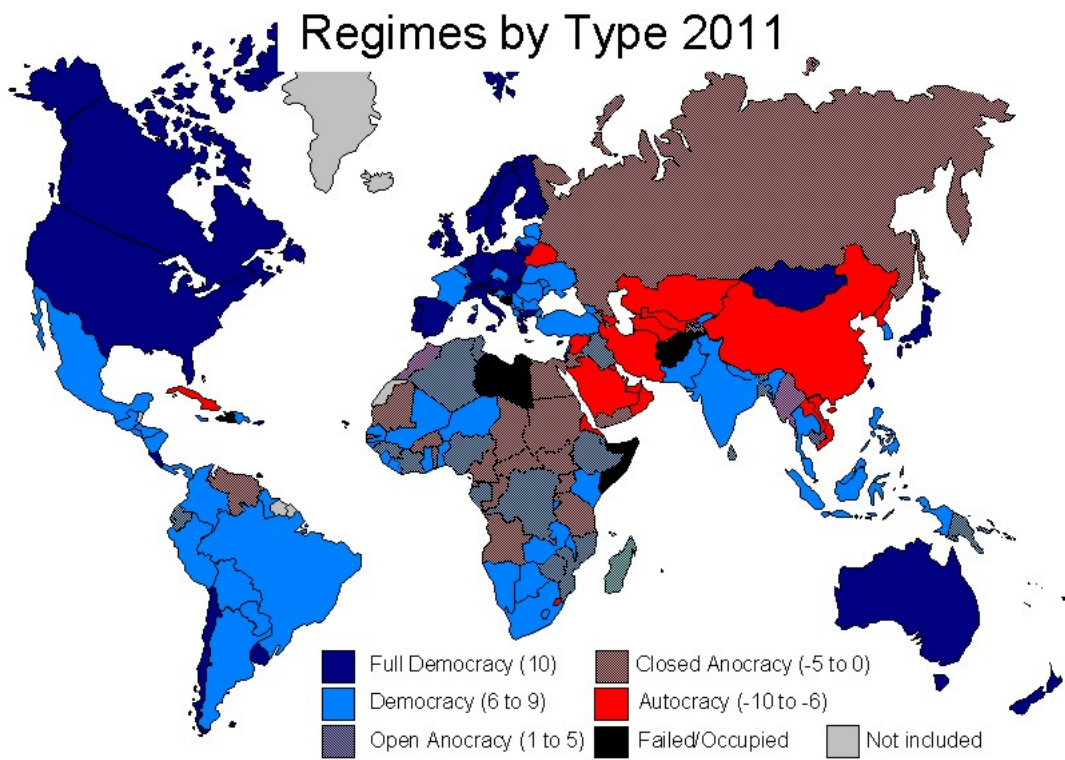
Figure 5: Log GNI per capita by regime type



Data from Quality of Government dataset

In my survey experiments, I test the two possible sources of confounding as mentioned above. (There are many other ways the naïve treatment vignette could cause confounding by making subjects think about other characteristics of the aggressor country.) To this end, I hypothesize the following:

Figure 6: Regime type map



Map from the Polity IV project

Hypothesis 1: Placebo tests will show that the naïve vignettes cause confounding by changing subjects' beliefs about the aggressor country's characteristics apart from its regime type.

One way to reduce confounding is to add in control variables, just as one would in an observation study:

Hypothesis 2: The vignettes with controls will cause less confounding than the naïve treatment by holding constant other characteristics of the aggressor country.

Nevertheless, one cannot control for all confounding factors, just as in an observational study. Furthermore, researchers who have limited funds cannot run multiple treatments with different controls. Therefore, I propose a more efficient alternative: embedding an invented "natural experiment" in the treatment vignette. Developing a plausible natural experiment embed requires some ingenuity, but one chief advantage is that it will make the difference between the control vignette and treatment vignette seem as-if random. Therefore I hypothesize the following:

Hypothesis 3: The natural experiment embed vignette will cause the least confounding out of the three types of vignettes.

In addition to testing for confounding, I am also interested in the fundamental question of whether Americans are more supportive of going to war against autocracies than democracies. Observing that all of the survey experiments' results have shown this to be true, I hypothesize the following:

Hypothesis 4: In all three survey experiments, each using a different vignette as described above, subjects will express higher support for military action against the dictatorship than against the democracy.

2.3.2 *Survey experiment methodology*

I conduct three sets of mini-survey experiments, each with a pair of treatment/control vignettes. Fielding these experiments on Amazon.com's Mechanical Turk, an online crowdsourcing service, I recruit 160 subjects for each mini-experiment. I acknowledge there are flaws associated with that web service. Recent research suggests that Mechanical Turk respondents, though less representative of the U.S. population than commercial survey panels, are often more representative than in-person convenience samples (Berinsky, Huber & Lenz 2012).

2.3.2.1 *The naïve treatment vignette experiment*

The first experiment consists of a simple, standard treatment of regime type. Subjects read a vignette about a country attacking its neighbor. For half of the subjects (randomly chosen), the aggressor country is a democracy, defined as a country with a democratically-elected government. For the other half of the subjects, the aggressor country is ruled by a dictator.³ After reading the vignettes, subjects are asked whether they support or oppose U.S. military action against the aggressor country.⁴ Next, I conduct placebo tests by asking readers 1) how likely or unlikely the conflict between the aggressor country and its neighbor took place in a economically developed part of the world and 2) what are the three plausible regions of the world the scenario could take place. Finally, subjects answer a series of questions about their demographic background, political identification, foreign language skills, and news consumption habits. Ideally, one would ask these questions pre-treatment, but I did not want to overload subjects' mental capacity before they process the treatment vignettes. I assume that the vignettes would not affect subjects' response to these background questions.

³ The exact vignette wording: "A country sends its military to take over a neighboring country. The attacking country is led by [a dictator/a democratically-elected government], who invades to get more power and resources." See Appendix A for the full question wording.

⁴ Subjects are asked "Do you agree or disagree that the U.S. should use its military to stop the aggressor country from invading its neighbor?" The answer choices are 1) strongly agree, 2) somewhat agree, 3) neither agree nor disagree, 4) somewhat disagree, 5) strongly disagree, 6) I don't know.

2.3.2.2 *The vignette-with-controls experiment*

In the second experiment, I control for possible confounds in the treatment vignettes. Subjects read a vignette about a country attacking its neighbor. The definitions of democracy and autocracy are identical to the ones in the the naïve vignette experiment. The aggressor country will be one of these four types:

	democracy	autocracy
developed economy and major U.S. trading partner	1	2
developing economy and minor U.S. trading partner	3	4

I use economic development and trade as controls because they present an opportunity to experimentally test Gartzke's (2007) "capitalist peace" theory. After reading the vignettes, subjects are asked whether they support or oppose U.S. military action against the aggressor country. Next, I conduct placebo tests by asking readers what they think the aggressor country is like in terms of other possible confounds (e.g., economic development⁵ and geographic region. Like the previous experiment, I also ask subjects a series of questions about their demographic background, political identification, foreign language skills, and news consumption habits.

2.3.2.3 *The "natural experiment" vignette experiment*

In the third experiment, I control for possible confounds by embedding a "natural experiment" that makes the difference in regime type between the treatment and control vignettes seem as if randomly. One can think of this design as a two-level randomization. Vignettes are randomly assigned to readers *and* the aggressor country's regime type, as depicted in the vignettes, is randomly assigned. Subjects read a vignette describing the recent history of the would-be aggressor country, labeled as Country A. All subjects read the following:

⁵ Although I control for economic development in the vignettes, I ask about economic development as a manipulation check.

Five years ago, Country A was a fragile democracy. It had a democratically elected government, headed by a popular president. At the time, a well-researched U.S. State Department report concluded that Country A was a democracy mainly because its president was very popular. The report also commented that without the president, there was a high probability that the country's military would overthrow the government to set up a dictatorship.

Half of subjects are randomly assigned to read the following (the democracy "treatment" condition):

Two years ago at a public event, a disgruntled military officer shot at the president of Country A. The president was hit in the shoulder and survived the attack. Country A's democratically elected government survived the political turmoil, and is still a democracy today.

Half of subjects are randomly assigned to read the following (the autocracy "control" condition):

Two years ago at a public event, a disgruntled military officer shot at the president of Country A. The president was hit in the head and did not survive the attack. In the political vacuum that followed the president's death, the country's military overthrew the democratically-elected government. Today, Country A is a military dictatorship.

Although these scenarios may seem strange, I was inspired by Jones and Olken's (2009) "hit or miss" natural experiment, which exploits the inherent randomness of assassination success/failure to study changes in political institutions.

After reading the vignettes about the background of Country A, readers are told this country invaded its neighbor. Subjects are asked whether they support or oppose U.S. military action against Country A. Like before, I conduct placebo tests by asking readers what they think the aggressor country is like in terms of economic development and geographic region. Finally, subjects answer a series of questions about their demographic background, political identification, foreign language skills, and news consumption habits.

2.4 RESULTS AND DISCUSSION

The results from a pilot study are inconsistent but interesting. They suggest that the treatment vignettes are not as confounded as I imagine. The naïve treatment vignette passed the placebo test question that detects economic development as a confounder. Nevertheless, the naïve vignette failed to pass some placebo tests designed to detect geographic location as a confounder. In contrast, the treatment vignette with controls and the natural experiment embed vignette passed all placebo tests. The substantive findings from my survey experiments are very different from results of previous studies. In the experiments with the naïve vignette and the vignette with controls, subjects were equally likely to support war against democracies and dictatorships. Only subjects in the experiment with the natural experiment embed vignette expressed significantly less support for going to war against democracies than dictatorships.

Before proceeding to the analysis of this survey experiment, I describe the data I have collected. Next, I discuss the implications of my data for the four hypotheses I seek to test.⁶

2.4.1 *Data*

The survey experiment was conducted in April 2013. Using Amazon.com's Mechanical Turk, I obtained 500 subjects. For the purpose of this analysis, I only consider subjects located in the U.S. and are American citizens. Given these conditions, I derive a final sample of 477 subjects. In the naïve vignette experiment, 75 subjects were assigned to the control group (dictatorship) and 85 were assigned to the treatment group (democracy). In the vignette with controls experiment, 90 were assigned to the control group (dictatorship) and 77 were assigned to the treatment group (democracy). In the natural experiment embed

⁶ My pre-analysis plan is available upon request.

vignette experiment, 66 were assigned to the control group (dictatorship) and 81 were assigned to the treatment group (democracy).

The dependent variable of interest is each subject's support for war against the aggressor country. I label this variable *support score*. The *support score* is derived from the question following the treatment vignettes. In the question, subjects are asked how much they agree or disagree that the U.S. should use military force against the aggressor country. I construct the *support score* in this manner:

- Strongly disagree: *support score*=1
- Somewhat disagree: *support score*=2
- Neither agree nor disagree: *support score*=3
- Somewhat agree: *support score*=4
- Strongly agree: *support score*=5

Thirteen subjects (2.74 percent of all respondents) selected "I don't know" as their answer. I cannot drop these responses because they are not independently distributed across treatment and control; pooling results from the three experiments, there were 11 "I don't knows" in the control group (dictatorship) and only two in the treatment group (democracy). So, I give each of the "I don't know" responses a *support score* of 3.

The first placebo test question subjects answered is the following: "How likely do you think it is that a scenario such as the one we just described could take place in an economically developed part of the world?" I code the answer choices in the following way:

- Very likely=1
- Somewhat likely=2
- Undecided=3
- Somewhat unlikely=4
- Very unlikely=5

For the second placebo test question, subjects are given a list of 11 geographic regions.⁷ Then subjects are asked: “Of all the regions in the world, what regions do you think are most likely to experience a scenario such as the one we just described?” Subjects pick the three most likely regions from the list of 11 regions.

The covariates I include for analysis are:

- age: continuous variable (minimum=18, maximum=73)
- sex: indicator variable (0=female, 1=male)
- education: indicator variable (0=no college degree, 1=college degree)
- party identification: indicator variable (0=Republican, 1=Democrat)⁸
- political ideology: 7-point scale (1=very liberal, 7=very conservative)
- foreign language(s): indicator variable (0=only speaks English; 1=speaks one or more foreign languages)

To check the random assignment procedure is sound, I check for covariates imbalance.⁹ As Table 1 shows, the pre-treatment covariates are essentially the same across treatment and control. Furthermore, when I regress the treatment assignment on these covariates, the beta coefficient for each is not significant. Thus, I conclude that my randomization procedure has worked and no administrative errors were committed.

2.4.2 Placebo tests

To test for confounding in the treatment vignettes, I compare treated subjects’ responses to the three placebo questions with control subjects’ responses. If treated subjects’ responses to the placebo questions are the same as control subjects’ responses for a given vignette pair, the vignette pair is not confounded. On the other hand, if treated subjects’ responses

⁷ The regions include North America, Central America and the Caribbean, South America, Western Europe, Eastern Europe, Middle East and North Africa, Sub-Saharan Africa, Central Asia, East Asia, South Asia, and Oceania.

⁸ I define Republican as subjects who label themselves as “strong Republican,” “weak Republican,” and “independent, leaning Republican.” Likewise, I define Democrats as subjects who label themselves as “strong Democrat,” “weak Democrat,” and “independent, leaning Democrat.”

⁹ The randomization is done using Qualtrics’s online software built-in randomization tool.

Table 1: Covariates balance check

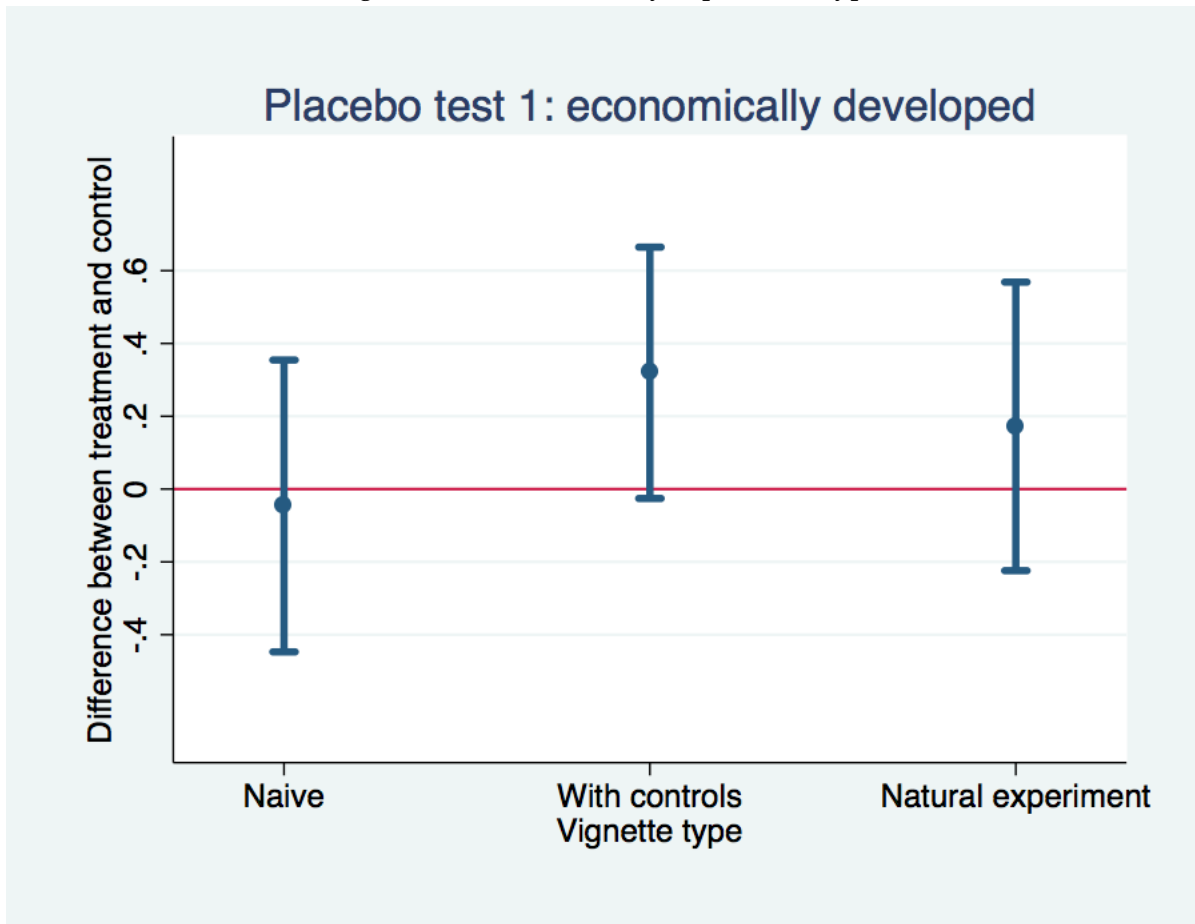
	Dictatorship	Democracy
Mean age	34	33
Mean ideology score	3.386957	3.345679
Proportion male	0.5130435	0.5596708
Proportion Democrat	0.5869565	0.5349794
Proportion white	0.8521739	0.8559671
Proportion speaks foreign language(s)	0.1913043	0.1934156

are different from control subjects', then the vignette is likely to be confounded. In the democratic peace survey experiment, for each vignette type, I test (using randomization inference, logit, and MANOVA) whether subjects' responses to the placebo questions are different across regime types. Figures 7 and Tables 2 and 3 showcase my findings.

For the first placebo question, subjects were asked how likely they think the treatment scenario would occur in an economically developed part of the world. All three vignette types (naïve, with controls, and natural experiment) passed the placebo test. As Figure 7 shows, there exists no difference between treated subjects' and control subjects' responses to the question for all three vignette types. Nevertheless, some vignette types did better than others. Contrary to my first hypothesis, the naïve treatment had the least amount of confounding since its difference in responses by regime types is closest to zero. The vignettes with controls almost failed the placebo test. In the vignettes with controls experiment, subjects think democracies are more likely to be economically developed than autocracies. This is particularly strange because the controls I introduce are supposed to prevent confounding about economic development. The natural experiment vignette performed fairly well in this placebo test and the difference in responses across regime types is minor. Overall, I cannot draw decisive conclusions because the sample sizes of these experiments are rather small.

For the second placebo test, subjects are given 11 regions of the world and are asked to pick the three most plausible regions where the scenario described in the vignette could take place. For each subjects' response, I coded a region 1 if he selects it as one of his plausi-

Figure 7: Placebo test 1 by experiment type



ble regions and coded a region 0 if he did not select it as one of his plausible regions. Then for each region of the world by vignette type, I perform logit regressions to test whether treated subjects and control subjects had different responses. My findings are reported in Table 2. (Oceania was omitted because too few subjects selected Oceania as a plausible region.) The vignettes with controls and the vignettes with the natural experiment embed passed all aspects of the placebo test; on average, subjects in both treatment and control selected the same plausible regions.

In contrast, the naïve vignettes did not pass two aspects of the placebo test. Subjects given the naïve vignette of democracy are significantly less likely to think the scenario could take place in the Middle East than subjects given the naïve vignette of dictatorship. This confounding is troublesome because 134 out of 160 subjects given the naïve vignette selected the Middle East as one of their plausible regions. Overall, most subjects selected the Middle East as one of their plausible regions. Because the Middle East is so salient in subjects' minds, researchers ought to ensure, at least, equal numbers of treated subjects and control subjects are thinking about the region. On the other hand, subjects given the naïve vignette of democracy are significantly more likely to think the scenario could take place in Central Asia than subjects given the naïve vignette of dictatorship. The confounding in the Central Asia case is not as troublesome because only 25 out of 160 subjects selected it as a plausible region.

After performing individual logit regressions, I study aggregate confounding across the three vignette types. For each of the three MANOVA, I examine subjects' responses to all 11 regions (0=not plausible; 1=plausible) by treatment and control. My findings are reported in Table 3. The naïve vignette experienced the most confounding because the differences between treated subjects and control subjects' response is the greatest. In fact, the difference in response between treated subjects and control subjects is almost statistically significant ($p=0.054$). On the other hand, the vignettes with controls and the vignettes with the natural experiment embed experienced no aggregate confounding. For those vignette types, there exists no statistical difference between responses from the treatment group and

control group. Nevertheless, contrary to my third hypothesis, the vignettes with controls fared slightly better than the vignettes with the natural experiment embed. The p-value of the former (0.773) is higher than the p-value of the latter (0.319). This difference though, is small, when one compares these two p-values with the p-value of the naïve vignette.¹⁰

My findings suggest that confounding vignettes pose a problem in survey experiments testing the democratic peace, but overall, confounding is a smaller problem than I imagined. One reason the naïve vignettes do not cause as much confounding as I hypothesized is that subjects have strong priors about the behavior of democracies and dictatorships. In the real world, the percent of democracies that invades neighboring countries is very small. So when subjects read about a democracy that invaded its neighbor, they might think the aggressor country is not a true democracy but one in name only.

I formalize the previous argument using Bayes' Theorem:

$$P(D|I) = \frac{P(I|D)P(D)}{P(I)} \quad (6)$$

where D is defined as Country X being a true democracy and I is defined as Country X invading its neighbor. P(I|D), the percent of true democracies that have invaded neighboring countries is fairly low. Let us set P(I|D) at 0.05. If we define true democracy D as countries with Quality of Government freedom scores of 8 or above, then P(D) is roughly around 0.5. Let us also define P(I|A) as the percent of non-true democracies that have invaded neighboring countries. I set P(I|A) higher than P(I|D), at 0.15, since this is observed in the real world that a greater percentage of autocracies than democracies fight with their neighbors. P(A) = 0.5 since P(A) is simply 1 – P(D). In this case,

$$P(D|I) = \frac{P(I|D)P(D)}{P(I)} = \frac{P(I|D)P(D)}{P(I|D)P(D) + P(I|A)P(A)} = \frac{(0.05)(0.5)}{(0.05)(0.5) + (0.15)(0.5)} = 0.25 \quad (7)$$

¹⁰ Given that the dependent variables I used were binary, MANOVA is far from ideal. A better alternative is to use is the nonparametric combination of dependent test (Caughey, Dafoe & Seawright 2013).

Table 2: Placebo tests: regions of the world
 Logit regressions
 Dependent variable: plausible region of the world (0=not plausible, 1=plausible)
 Independent variable: regime type (0=dictatorship, 1=democracy)

	Naïve			With controls			Natural experiment		
	Coef.	Robust SE	P-value	Coef.	Robust SE	P-value	Coef.	Robust SE	P-value
Subsaharan Africa	-0.373	0.328	0.254	-0.019	0.314	0.951	0.058	0.339	0.863
Middle East	-1.525	0.528	0.004	-0.390	0.461	0.398	-0.191	0.471	0.685
East Asia	0.486	0.368	0.188	-0.354	0.344	0.303	0.387	0.390	0.321
South America	0.261	0.356	0.463	-0.372	0.349	0.285	0.195	0.357	0.585
Western Europe	0.061	0.629	0.922	0.057	0.490	0.907	0.744	0.857	0.385
Central Asia	0.959	0.479	0.045	0.055	0.338	0.871	-0.076	0.423	0.857
North America	1.531	1.111	0.168	0.700	0.750	0.350	1.217	1.135	0.283
Central America	-0.254	0.470	0.589	0.453	0.424	0.286	-0.806	0.486	0.097
Eastern Europe	-0.235	0.402	0.558	0.353	0.406	0.384	0.326	0.485	0.501
South Asia	-0.201	0.328	0.541	-0.080	0.347	0.817	0.109	0.350	0.755
N	160			167			147		

Table 3: Placebo test: regions of the world
MANOVA

Dependent variables: 11 regions of the world (0=not plausible, 1=plausible)
Independent variable: regime type (0=dictatorship, 1=democracy)

	Naïve			With controls			Natural experiment		
	Statistic	F	P-value	Statistic	F	P-value	Statistic	F	P-value
Wilks' lambda	0.880	1.830	0.054	0.955	0.660	0.773	0.914	1.160	0.319
Pillai's trace	0.120	1.830	0.054	0.045	0.660	0.773	0.087	1.160	0.319
Lawley-Hotelling trace	0.136	1.830	0.054	0.047	0.660	0.773	0.095	1.160	0.319
Roy's largest root	0.136	1.830	0.054	0.047	0.660	0.773	0.095	1.160	0.319
N	160			167			147		

If the probability of a country is a true democracy given it has invaded its neighbor is 0.25 or some other low number, subjects might be wary of the naïve vignettes. Although I label the aggressor country as a democracy, subjects might not believe my vignette. They might think the aggressor country is a fragile democracy with the stereotypical characteristics of a dictatorship, such as the lack of economic development and location in a poor region of the world. In simply put, subjects treated with the naïve treatment vignette of democracy are not thinking about Sweden but are thinking about South Sudan.

If subjects in survey experiments are updating in a Bayesian way, then it would be difficult to test “big theories” using vignettes. Vignettes require extreme counterfactuals to test many of the big theories in political science. But if subjects have strong priors about how the world operates, implausible counterfactuals might not move their beliefs. This is perhaps equally as troubling as the problem of confounding caused by poorly designed treatment vignettes.

2.4.3 *Substantive findings*

Aside from the placebo tests, I also estimated the ATEs of democracy on subjects’ support for war across the three vignette types. I define ATE as $E[Y_i(\text{Democracy}) - Y_i(\text{Dictatorship})]$ and estimate it for each vignette type using randomization inference and linear regression. Figure 8 and Table 4 and 5 report my findings. Whether I estimate the ATEs using randomization inference or linear regression, the effect of the aggressor being a democracy is not significant for the naïve vignette type or the vignette with controls type. Only the vignettes with the natural experiment embed produced significant effects: the treatment of democracy lowered *support scores* by more than 0.5 points — as consistent with the theory of democratic peace.

Why did the naïve vignette and the vignette with controls fail to move subjects? One possible reason could be that vignettes in them are rather bland and did not engage the reader’s attention. Readers might have not paid attention to the part of the vignette that

mentioned the aggressor country's regime type because I did not bold or highlight it.¹¹ In contrast, the natural experiment vignettes contained rich narratives and focused on the political institution of the aggressor country. Therefore, subjects might have thought more about the regime type of the aggressor country when they are asked whether they agree or disagree about going to war.

It would be irresponsible for me to conclude that my survey experiment failed to replicate the results of previous survey experiments conducted about this subject. While Mechanical Turk sample might be more representative than a convenience sample of college students, it is far worse than a sample from an online survey service like YouGov or Knowledge Networks used by Tomz and Weeks for their experiments. Furthermore, the sample size for each of the experiments I ran is fairly small, so my experiments might be under-powered. In the future, when implementing the real version of this survey experiment, I will use a nationally representative sample and write more engaging vignettes that highlight the institutional differences between democracies and autocracies.

Table 4: Estimating the ATE of democracy: without pre-treatment covariates

		Coef.	Robust SE	t	P-value
Naïve N=160	democracy	-0.105	0.171	-0.610	0.540
	constant	3.093	0.122	25.410	0.000
With controls N=167	democracy	-0.041	0.172	-0.240	0.813
	constant	3.067	0.120	25.520	0.000
Natural experiment N = 147	democracy	-0.559	0.172	-3.250	0.001
	constant	3.152	0.135	23.320	0.000

¹¹ I did so deliberately because I did not want to make subjects conscious of the treatment.

Figure 8: ATEs by experiment type
 Estimated using randomization inference; 95 percent confidence intervals

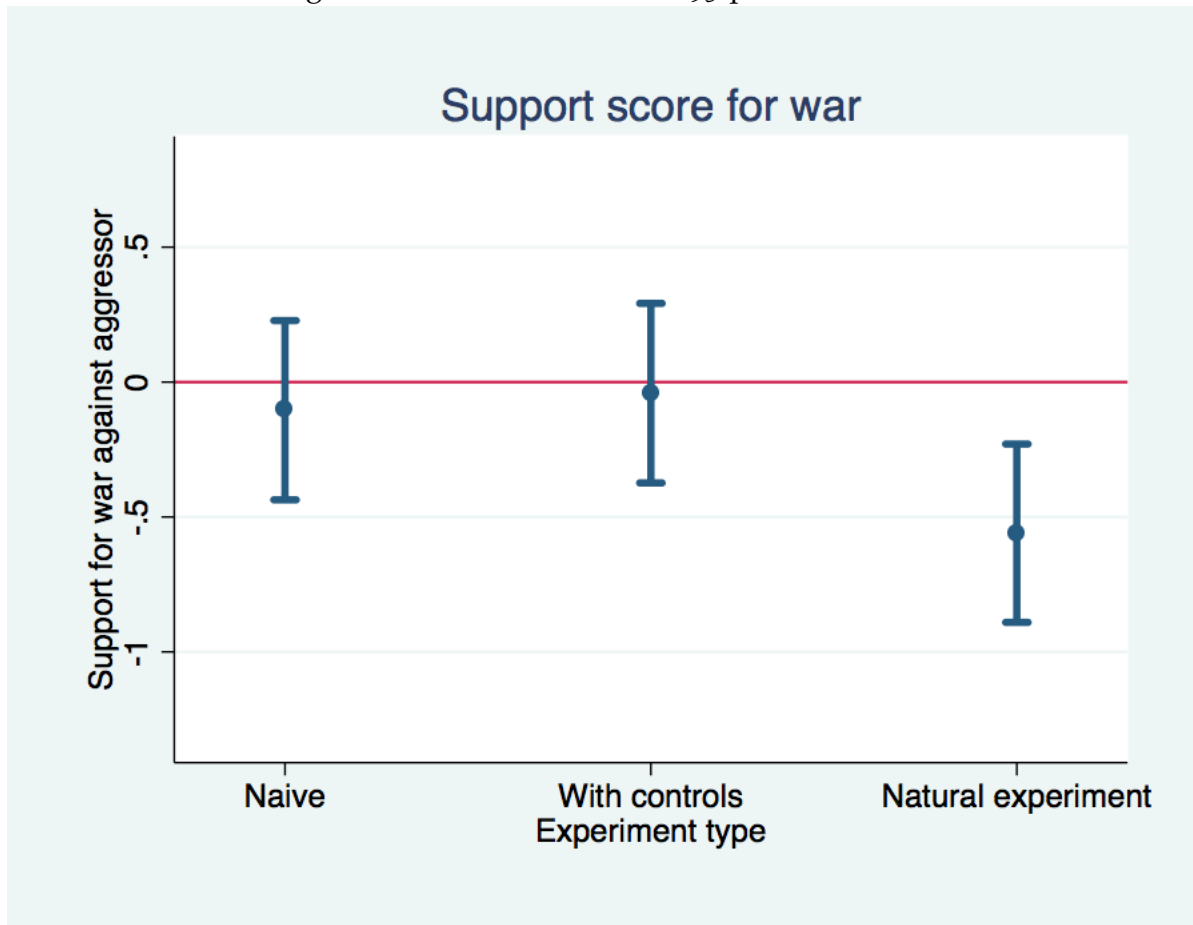


Table 5: Estimating the ATE of democracy: with pre-treatment covariates

		Coef.	Robust SE	t	P-value
Naïve N=160	democracy	-0.092	0.170	-0.540	0.591
	constant	3.234	0.443	7.300	0.000
With controls N=167	democracy	-0.060	0.176	-0.340	0.735
	constant	2.914	0.562	5.180	0.000
Natural experiment N = 147	democracy	-0.534	0.170	-3.140	0.002
	constant	3.434	0.523	6.560	0.000

2.5 CONCLUSION

The three survey experiments I conducted seek to test the theory of democratic peace. Results from these experiments, in part, support the theory that citizens in democracies are less supportive of launching wars against other democracies than autocracies. When the difference in regime type is made salient, as in the vignette with the natural experiment embed, subjects were less likely to support using military action against the democracy than the dictatorship. But when the difference in regime types is not made salient, as in the other two vignette types, subjects expressed equal support for military action against both types of aggressor countries. Reflecting upon the external validity of this survey experiment, I argue the vignette with the natural experiment embed is most like the transfer of news information in the real world. American political leaders and news sources often highlight the political institutions of aggressor countries that might come into conflict with the U.S., especially if those countries are non-democratic.

Results from the placebo tests suggest that the naïve vignette previously used in survey experiments on this subject might not be as confounded as I suspected. Although democracies, on average, are wealthier than autocracies, subjects do not consider the democratic aggressor to be more economically developed than the autocratic aggressor. On one hand, the lack of confounding along this dimension is good because researchers can be confident when drawing inference. But the results also suggest that subjects are considering a small subset of countries in the world when reading the treatment vignettes. Because subjects have strong priors about how the world operates (e.g., Sweden will not invade Norway), they are likely to think about fragile democracies that almost operate as autocracies when they are told about a democracy that invaded its neighbor. If this is the case, then it would be difficult for IR scholars to understand why European countries no longer go to war against each other because subjects' minds are instead fixated on developing countries in Sub-Saharan Africa or the Middle East.

If survey subjects are indeed updating in a Bayesian way, then experimenters should think carefully about design. Instead of testing extreme counterfactuals just because they are dictated by theory, researchers should introduce plausible scenarios in vignettes. Implausible scenarios may cause subjects to disbelieve the treatment vignette and lead them to answer questions in unintended ways. The obvious downside to plausible scenarios is that they reduce big political theories to a subset of cases. But such is the trade-off when one attempts to draw unbiased inference from survey experiments.

2.6 ACKNOWLEDGMENTS

This chapter is part of a joint project with Allan Dafoe, who originally provided me with the idea of confounding vignettes in survey experiments, and Devin Caughey. I thank Alan Gerber for helping me formalize the problem within a structural equation modeling framework. I am also grateful for Peter Aronow, Donald Green, and Jas Sekhon's advice on methodology and John Bullock's advice on public opinion research

3 UN ENDORSEMENTS' EFFECT ON PUBLIC OPINION

Abstract. Recent scholarship on international institutions suggest organizations, such as the United Nations (UN), influence states' domestic public opinion about military interventions in foreign countries. Endorsements by international organizations serve as elite cues for domestic audiences. In this paper, I provide evidence from a survey experiment that demonstrates UN approval increases public support for military intervention and economic sanctions in response to both humanitarian crisis and security threats. Conditional average treatment effect is especially large for subjects who strongly support the UN. To assess the external validity of this and similar survey experiments, I tested whether subjects would respond differently to treatment prompts with no country names (generic prompts) compared to those with country names (specific prompts). A difference-in-difference analysis suggests that the treatment effect of UN approval is similar across generic prompts and specific prompts.

3.1 INTRODUCTION

Do international organizations (IOs) influence domestic politics? Traditionally, scholars had ignored the role of IOs in the intersection between international relations and domestic politics; instead, they focused on national leaders and diplomats. Leaders of democratic countries worry about domestic audience cost when deciding to take military action against another state (Fearon 1994). Diplomats in international negotiations engage in a two-level game with other diplomats and with their domestic government (Putnam 1988).

But a more recent line of research centers on the indirect impact of IOs on domestic public opinion. In democratic countries, the attitudes of the mass public constrain leaders to behave in certain ways. For instance, when a president or prime minister tries to persuade the public to go to war, the public can get a “second opinion” from the United Nations Security Council. In effect, authorization by the UN Security Council legitimizes military interventions in the eyes of the masses. Chapman and Reiter’s (2004) study of historical polling data and Grieco et al. (2011) and Tingley and Tomz’s (2012) survey experiments contend that the UN affect mass public support for war.

I contribute to this small but growing literature in two ways using a survey experiment. First, I expand upon previous studies to consider whether UN endorsements also influence public opinion about economic sanctions, something that has never been tested experimentally. Second, I determine whether UN endorsement’s impact on public opinion vary as the motivations behind the interventions vary. In the survey experiment, subjects were presented with scenario prompts that contained hypothetical cases of international crises (a nuclear security threat or a genocide) followed by potential intervention measures (military action or economic sanctions). Subjects were randomly sorted into two groups: the treatment group received prompts in which the UN endorsed the intervention measures; the control group received prompts in which the UN disapproved of the intervention measures. I find that endorsements by the UN increase public support for military action and economic sanctions in response to both humanitarian crisis and national security threats.

UN approval has approximately the same effect on subjects' attitudes towards the intervention measures, whether it is to deter a nuclear security threat or to stop genocide. The treatment effect of UN endorsement is significantly greater for subjects who support the UN than for those who hardly value the UN. To assess the external validity of this and similar survey experiments, I test whether subjects would respond differently to scenario prompts with no country names (generic prompts) than to those with country names (specific prompts). A difference-in-difference analysis suggests that the treatment effect of UN approval is similar across generic prompts and specific prompts.

3.2 LITERATURE REVIEW

There exists much theoretical work about why states consult formal IOs before engaging in international conflicts. It is widely assumed that because IOs hold a degree of independence from other states, they provide the most unbiased information. Therefore, by obtaining approval from the UN Security Council, a state can cast its military actions – even unilateral ones – as legitimate (Abbott & Snidal 1998). Voeten (2005) expands upon this theory by arguing the Security Council's authorizations serve as an information shortcut for the domestic audience. Americans generally want their country to be involved internationally, but they also fear overextension of the U.S. military. An authorization by the UN indicates that no costly challenges will result from the military action.

This theory of how IOs influence public opinion about military interventions has support in empirical evidence from a study of historical data and two survey experiments. Chapman and Reiter (2004) examine polling data to see how Americans responded to involvements in militarized interstate disputes between 1945 and 2001. They note that presidential approval ratings were 9 percentage points higher, on average, in missions that had UN Security Council approval. Although Chapman and Rieter find a positive correlation between UN approval and public support for the president (implicitly the military intervention), one cannot conclude that the former caused the latter. For instance,

reverse causality is possible: historically, the Security Council might have simply approved the interventions that everyone, including the mass public, agrees is legitimate. Grieco et al. (2011) and Tingley and Tomz's (2012) survey experiments overcome the limitations of Chapman and Reiter's observational study.

The data for Grieco et al.'s (2011) paper came from a 2003 survey experiment conducted over the phone. The subjects were presented with a prompt that described the hypothetical invasion of East Timor by Indonesia. The treatment group received a cue indicating the UN and NATO supported a U.S. military mission in East Timor. In contrast, the control group received a cue that stated the UN and NATO does not approve of such an operation. True to the authors' hypothesis, the IO cue has a significant effect on increasing individual subjects' support for military intervention in East Timor. The effect is most notable among those who valued the UN and NATO and those who had little confidence in the president. Grieco et al. admit the shortcoming of their survey experiment's external validity: the prompt they presented is representative only of low-cost humanitarian interventions.

Tingley and Tomz's (2012) study furthers the discussion by testing why UN Security Council authorization affects public opinion. They hypothesize three reasons: first, the UN cue suggests military force is warranted; second, the UN cue indicates that other countries will share the military burden; third, the UN cue signals a public commitment to contribute to a multinational effort. The third reason is the most consistent with the data from the authors' survey experiment.

This paper, based on a pilot study using Yale University undergraduates, builds upon the papers discussed above. Chiefly, I test whether the effect of UN approval varies across different types of intervention measures. In a future survey experiment I plan to conduct on Amazon.com's Mechanical Turk, I hope expand upon Tingley and Tomz's work by testing causal mechanisms behind the effect of UN approval.

3.3 HYPOTHESES AND RESEARCH DESIGN

This section lays out the methodology behind my study: first, I outline my four hypotheses; second, I explain the research design of my survey experiment; third, I describe how I collected my data.

3.3.1 *Hypotheses*

The four hypotheses of my study are as stated:

Hypothesis 1: UN approval increases public support for military interventions and economic sanctions in response to both national security threats and humanitarian crises.

While there exists much literature about public opinion and American presidents' decisions to launch wars, scholars have largely ignored the relationship between public opinion and economic sanctions. Drury's (2001) study, using historical data, suggests that presidents were more likely to issue economic sanctions when his approval ratings were high and when unemployment rates were low. While Drury acknowledges that public opinion has marginal effects on presidents' decisions, he asserts that unpopular presidents nearing elections are unlikely to issue economic sanctions (Drury 2005). I argue that while war costs much more than economic sanctions, the American public might feel at least some cost. In the worst case, the price of gasoline or other highly demanded commodity might dramatically increase. Therefore, UN legitimization for economic sanctions provides the assurance to Americans that the intervention measure is warranted.

Hypothesis 2: The effect of UN approval on public support for military interventions and economic sanctions is greater in response to humanitarian crises than to national security threats.

I arrived Hypothesis 2 because we can assume that Americans value the preservation of their state. If there exists a grave, existential threat to the U.S., Americans are likely to support intervention measures – UN approval or not – to protect the security of their state. In contrast, military interventions and economic sanctions in response to humanitarian crises (specifically genocide in the survey experiment) are often elective measures. Although genocides are horrific, most of them are contained within a state, and therefore do not pose as grave threats to American national security. As a result, the American public has more freedom in choosing how to respond. Therefore, UN approval might have a greater effect on public support for intervention measures in response to humanitarian crises than to security threats.

Hypothesis 3: The conditional average treatment effect (CATEs) of UN endorsement are higher among Democrats and isolationists. CATEs are also higher for those who have little confidence in the Executive Branch or those who strongly support the UN.

In Grieco et al.'s (2011) study, the authors examine the effect of UN endorsements for different subsets of their sample. They discover people who have low confidence in the president experience and those who valued IOs experienced greater CATE. Unfortunately for the authors, the survey experiment was conducted on the eve of the U.S.'s invasion of Iraq. The lead up to the invasion pitted Americans who support Bush's decision to launch a preemptive strike against Iraq against those who distrusted the presidents' judgment. Furthermore, the anti-war public pointed to the lack of a second UN vote as evidence that the war was illegitimate. Therefore, the discourses of the times might have influenced survey respondents in an unusual way. To confirm the external validity of Grieco et al.'s survey experiment, I will try to replicate their findings.

In addition, I argue that Democrats and isolationists will experience larger CATEs. Since the 1960s, Democrats, more than Republicans, have become the party of cosmopolitans and cooperative internationalists (Gerring 2001). Furthermore, Hayes and Guardino's (2011) research have shown that Democrats, more than Republicans, were influenced by foreign

media during the lead up to the Iraq War. Non-isolationists would experience a small CATE from UN Endorsement because they are likely to support intervention either way. In contrast, UN endorsements might matter more to people who are perennially against interventions.

Hypothesis 4: Including country names in the treatment/control prompts does not change the average treatment effect of UN approval.

Grieco et al. (2011) and Tingley and Tomz's (2012) survey experiments employ different scenario prompts. Grieco et al. used specific prompts that characterized potential intervention in one country only. In contrast, Tingley and Tomz used generic prompts that stated the intervention would take place in "a country in Africa." Rather than choose one method over another, I decided to test both. In the specific prompts, I used a list of randomly, rotating country names. In the generic prompts, I used the term "a country" as the target for potential intervention. I hypothesize that adding country names in the prompts will not change the ATE of UN approval. Even when given generic prompts and told not to think about specific countries, subjects are likely to think about specific countries. This phenomena can be explained by ironic process theory, the psychological process whereby an individual's deliberate attempts to suppress or avoid certain thoughts render those thoughts more persistent (Wegner 1994). Therefore, the ATE of UN approval will be similar across generic prompts and specific prompts.

3.3.2 *Research design*

To test whether sanctions by the United Nations Security Council (UNSC) shape public opinion, I designed a new experiment and embedded in an online public opinion survey. The first part of the survey asked each subject about his or her basic demographic information (age, sex, and ethnicity), partisanship, and political ideology. The demographic questions were included for balance tests and to improve the precision of regression analysis. Next, the survey asked each subject his/her opinions on the executive branch of the

U.S. government, military spending, the UN, and human rights. I included those questions in the survey to test whether subsets of subjects are more likely to be swayed by the treatment effect (i.e., UN endorsement of military or economic intervention).

Before moving onto the experimental part of the survey, each subject read instructions on how to answer the questions that follow. Adapted from Tomz and Weeks's (2012a) survey experiment about democratic peace, the instructions read:

You will read about situations our country has faced many times in the past and will probably face again. Some of the situations are general, and some of the situations are about specific countries. Some parts of each description may strike you as important; other parts may seem unimportant. After describing each situation, we will ask your opinion about a policy option.

Subjects were required to click "I understand the instructions" before moving onto the next part of the survey. Each subject was given four hypothetical scenario prompts that each dealt with an international crisis.

The experiment had a $4 \times 2 \times 2$ design. There were four categories of scenarios: 1) a security threat followed by potential military intervention, 2) a security threat followed by potential economic sanctions, 3) a humanitarian crisis followed by potential military intervention, and 4) a humanitarian crisis followed by potential economic sanctions. Within each category, there were four possible combinations of treatments and controls. See Table 6 for the four possible experimental groups.

To test whether the UN approval affects public opinion, I set the control to be no UN approval for potential interventions and the treatment to be UN approval for potential interventions. Furthermore, to test whether country names have an effect on responses, I set the control to be scenarios without country names and the treatment to be scenarios with country names. The treatment country names in the humanitarian crisis scenarios included Chad, Angola, Sudan, South Sudan, Burma, Burundi, Laos, Uganda, the Democratic Republic of the Congo, and the Central African Republic.¹ The treatment country names in the security threat scenarios were North Korea and Iran.

¹ These 10 countries experienced the most violence against ethnic minorities in 2005 according to the latest edition of the Minorities at Risk (MAR) dataset.

Table 6: Treatment/control chart

	Control: No UN endorsement	Treatment: UN endorsement
Control: No country name	0,0	0,1
Treatment: Country name	1,0	1,1

Each subject was presented with exactly one scenario prompt from each of the four categories; therefore, each subject was presented with four prompts in total. The order of those four prompts was randomized. For each prompt, the subject randomly received one of the four treatments-controls combinations. The tree flowchart in Figure 9 serves as visual explanation of the experiment's design: each subject reached four of the 16 pink nodes, given that he/she passed through each the blue nodes exactly once.

After reading each scenario, subjects were asked how much they oppose or favor the potential military intervention or economic sanctions. Subjects moved a slider along an 11-point scale to indicate their feeling towards each potential intervention measure. A 0 on the scale means one strongly opposes the measure. A 10 on the scale means one strongly supports the measure. These answer choices were the outcome measures of the experiment. Below are the possible scenarios that subjects were presented with:

Security Threat: [A country/"Named Country"] is developing nuclear weapons and will have its first nuclear bomb within six months. The country could then use its missiles to launch nuclear attacks against any country in the world, including the U.S. and its allies. [A country/"Named Country"] has refused all requests to stop its nuclear weapons program.

Democrats and Republicans in Congress have approved [the use of military force/economic sanctions], in order to prevent it from making any nuclear weapons. The UN has [not approved/also approved] of such a measure.

Question: How much do you oppose or favor U.S. [military action/economic sanctions] against [this country/"Named Country"]?

Indicate using the slider along the 11-point scale.

Humanitarian Crisis: The leaders of [a country/"Named Country"] have launched a policy to kill ethnic minorities in that country, with the intent to destroy them. The leaders of the country have refused all requests to stop the killings.

Figure 9: Treatment tree



Democrats and Republicans in Congress have approved [the use of military force/economic sanctions], to prevent or end the grave human rights violations taking place in [the country/“Named Country”]. The UN has [not approved/also approved] of such a measure.

Question: How much do you oppose or favor U.S. [military action/economic sanctions] against [this country/“Named Country”]?

Indicate using the slider along the 11-point scale.

Instead of using prompts that are fully abstract (e.g., “A country invaded another country”), I used scenarios that are general enough for subjects to understand but defined enough to fit within categories (security threat vs. humanitarian crisis). Unlike most experimental IR scholars, who tend to use militarized interstate disputes in their scenario prompts, I decided to use an intrastate conflict in my humanitarian crisis scenario. I consciously made this decision to make my scenarios mirror reality as much as possible: since 1990, civil wars has made up an overwhelming majority of armed conflicts.²

I fielded this experiment in a public opinion survey sent to Yale undergraduates in the class of 2016, 2015, and 2013 between September 19 and October 1, 2012. The experiment yielded a total of 1,108 respondents who are American citizens.

3.4 RESULTS AND DISCUSSION

The results of my survey experiment, for a large part, are consistent with my hypotheses. As expected, UN endorsements increase public support for military interventions and economic sanctions in response to both security threats and humanitarian crises. But unexpectedly, UN approval does not have a larger treatment effect in increasing public support for humanitarian interventions compared with security interventions. While people who have low confidence in the president do not experience higher CATE, subjects who strongly support the UN do. Finally, including country names in the scenario prompts do not change the ATE of UN approval on support for any intervention measure.

² See the Uppsala Conflict Data Program Armed Conflict Dataset.

3.4.1 *UN approval matters*

Before considering the results of the survey experiment, I first test its internal validity. First, I measured the relationship between treatment group assignment and the pre-treatment covariates. As Table 7 illustrates, the demographics and attitudes of the subjects in the UN endorsement treatment and control groups are nearly identical. In addition, subjects treated with specific prompts have the same background characteristics as those treated with generic prompts. Furthermore, when I performed an OLS regression using treatment assignment as the dependent variable and the pre-treatment covariates as the independent variables, none of the regressors were significant. Therefore, we can conclude random assignment has been achieved.

Table 7: Covariates balance check

	Control: No UN Endorsement	Treatment: UN Endorsement Treatment
Mean Age	18.980	18.999
Percentage Male	0.515	0.505
Percentage White	0.557	0.552
Percentage Republican	0.113	0.126
Mean Ideology (0=very liberal, 1=very conservative)	0.295	0.312
Mean Support for Human Rights (0=low support, 1=high support)	0.752	0.750
Mean Government Spending Score (0=too little spending, 1=too much spending)	0.842	0.853
Mean Confidence in Executive (0=little confidence, 1=great confidence)	0.568	0.557
Mean Isolationism Score (0=weak isolationism, 1=strong isolationism)	0.429	0.435
Mean UN Evaluation Score (0=poor job, 1=good job)	0.446	0.430
Mean Support for the UN (0=low support, 1=high support)	0.676	0.673
Mean Belief About UN Legitimization (0=weak belief, 1=strong belief)	0.640	0.643
N of observations	2934	2934

As consistent with Hypothesis 1, Figures 10 and Table 8 demonstrate that UN endorsement significantly increases public support for military intervention and economic sanctions in response to both security threats and humanitarian crises. Table 8 contains the difference-in-means between the treatment and control groups for both military action and economic sanctions, together and separately. I conclude that all three difference-in-means are significant after running t-tests for each. Furthermore, I conducted randomization tests using the R randomization inference package *ri* to estimate the ATE of each treatment. Samii and Aronow (2012) demonstrate that randomization-based constant effects estimator is equivalent to the OLS estimator using White’s “robust” standard errors. Overall, the UN endorsement treatment increased support for the intervention measures by 0.86 points. The estimated ATE for military action (1.06) is slightly larger than the estimated ATE for economic sanction (0.78).

Figure 10: Support for intervention measures

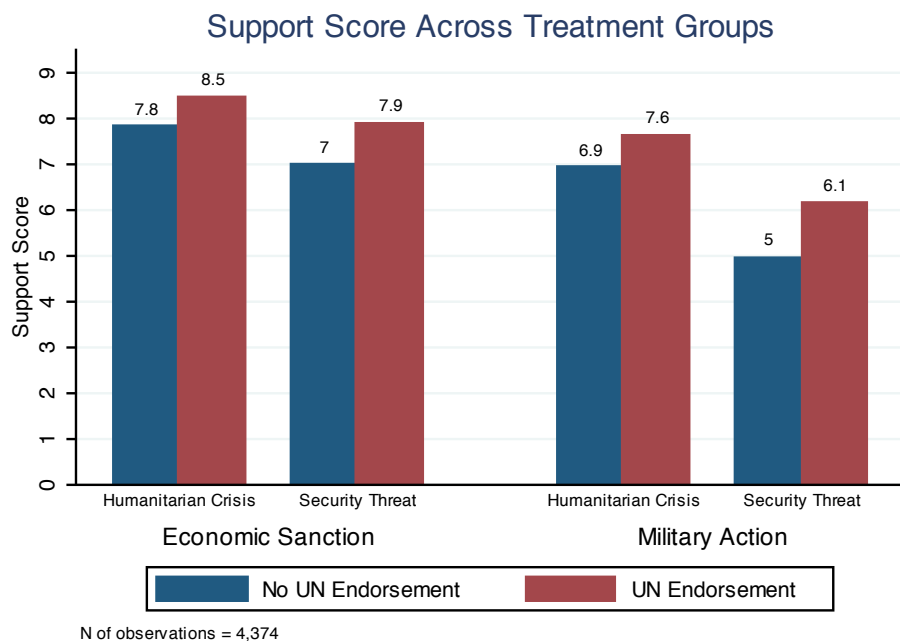


Table 8: Estimating the treatment effects of UN endorsements

	Overall	Military action	Economic sanctions
Difference-in-means	0.861	1.058	0.782
ATE estimated using ri	0.858	1.056	0.779
N of observations	4374	2189	2191

Due to the unusual nature of our experimental data, randomization tests might not accurately estimate the ATE of UN endorsements. Since each subject received four treatments, the observations are not independent from each other. To control for confounding effects of each individual subject and the order the four treatments, I constructed a panel dataset. The panel identification variable is each subject's unique ID number and the time variable is the order of the treatments.³ To analyze the effect of UN endorsement, I performed a random effects estimator (i.e., feasible GLS regression).

I begin with the unobserved effects model:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it} \quad (8)$$

where y_{it} is each support score, β_1 is the ATE of UN endorsement, x_{it} is the randomly assigned treatment, and α_i is the unobserved individual effect. In the random effects model, one must assume the unobserved effect α_i is uncorrelated with the explanatory variable in all time periods:

$$\text{Cov}(x_{itj}, \alpha_i) = 0, t = 1, 2, \dots, T; j = 1, 2, \dots, k. \quad (9)$$

This assumption holds because each subjects were randomly assigned treatments. When I define the composite error term as $v_{it} = \alpha_i + u_{it}$, I write model one as

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it} \quad (10)$$

³ I constructed a variable for the order of the treatments using the survey's meta-data of when subjects answered each question.

where $\text{Corr}(v_{it}, v_{is}) = \sigma_a^2 / (\sigma_a^2 + \sigma_u^2)$, $t \neq s$. σ_a^2 is the variance of a_i and σ_u^2 is the variance of the u_{it} . Because a_i is in the composite error in each time period, the v_{it} are serially correlated across time. Thus, I used GLS to solve the serial correlation problem. In order for the procedure to work, the dataset must have large N and relatively small T (Wooldridge 2001). My data fit both of these requirements. First I define

$$\lambda = 1 - [\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)]^{1/2} \quad (11)$$

. Next, I derive the transformed equation:

$$y_{it} - \lambda \bar{y}_i = \beta_0(1 - \lambda) + \beta_1(x_{it} - \lambda \bar{x}_{i1}) + (v_{it} - \lambda \bar{v}_i) \quad (12)$$

Because one can only use the feasible GLS estimator, one must substitute $\hat{\lambda}$ for λ . I used the random effects estimator instead of the fixed effects estimator because a_i , the individual effects, are uncorrelated with the treatment assignments. Furthermore, I performed a Hausman test that resulted in a Chi-square statistic of 0.75, which suggests the idiosyncratic errors and explanatory variable are uncorrelated across all time periods.

The results of the GLS regressions are reported in Table 9. The ATEs calculated are fairly similar to those calculated using randomization inference, although those for military action derived through regression analysis are somewhat lower. In Models 4 and 5, I included the covariates from pre-treatment questions for increased precision. The covariates included information about each subject's demographics, partisanship, and opinions about foreign policy and the UN. To make data interpretation easier, I standardized these covariates to range from 0 to 1. Including the covariates slightly increased the ATE of UN endorsements. Again, the results are similar to those found in Table 8.

In sum, UN endorsements significantly increase public support for military actions and economic sanctions. When there is UN endorsement, overall support increases by about 15

Table 9: GLS Regression: Random Effects

	Model 1	Model 2		
UN Endorsement	0.857 (0.074)***	0.749 (0.092)***		
Intervention Type	Military Action and Economic Sanction	Economic Sanction		
Constant	6.664 (0.062)***	7.402 (0.074)***		
N of Observations	4374	2191		
N of Subjects	1108	1103		
Covariates	No	No		
	Model 3	Model 4	Model 5	
UN Endorsement	0.926 (0.112)***	0.836 (0.112)***	0.973 (0.134)***	
Intervention Type	Military Action	Economic Sanctions	Military Action	
Constant	5.947 (0.081)***	7.925 (1.553)***	9.361 (1.607)***	
N of Observations	2183	1473	1462	
N of Subjects	1101	740	736	
Covariates	No	Yes	Yes	

Dependent variable: support score for intervention measure (0=strongly oppose, 10=strongly favor)
 Covariates include age, male, white, Republican, ideology, support for human rights, opinion about government spending, confidence in the executive branch of the government, level of isolationism, evaluation of the UN, support for the UN, and beliefs about UN legitimization

percent; support for military action increases by about 18 percent and economic sanctions about 11 percent.⁴

3.4.2 *Different intervention measures, different ATEs?*

Are the ATEs of UN endorsements the same across all types of interventions? Initially, the results from the difference-in-difference analysis in Figure 10 suggests the answer is no. In Model 7, I created an interaction term for UN endorsement and whether the intervention addresses a security threat. When I regressed the support score on this interaction term, the coefficient was positive and significant. This suggests that the ATE for security threats is 0.38 points greater than the ATE for humanitarian crises. The empirical evidence seems to contradict my Hypothesis 2, which suggests the opposite.

But before I reject my Hypothesis 2 and conclude that UN endorsements matter more for security interventions than humanitarian interventions, I must consider whether this difference in ATEs resulted from functional form. Given a treatment, one could assume that the ATE is constant across the support score levels. That means one assumes the UN endorsement treatment will raise someone's support level, on average, the same amount – no matter what his baseline support for the intervention is. (Baseline is defined as the support score in the control scenario of no UN endorsement.)

This assumption, however, may not be true. The ATE might have a logistic distribution. If this is the case, the effect of treatment is greatest for a baseline level in the middle (at 5), and weakest at the extremes (0 and 10). This functional form problem poses a threat to our interpretation of the data in Table 10 because interventions for humanitarian reasons have higher baseline support than interventions for security reasons. The bigger ATE one observes for the security interventions would have resulted from the functional form, not from the treatment assignment of intervention type.

⁴ These figures were calculated by performing the GLS regressions on the natural log of support scores, without including covariates.

Table 10: GLS regression: random effects; by intervention types

	Model 6	Model 7
UN Endorsement Treatment	0.851 (0.105)***	0.668 (0.101)***
Country Named	-0.075 (0.104)	
UN Endorsement*Country Named	0.012 (0.148)	
Security Threat		-1.403 (0.096)***
UN Endorsement*Security Treat		0.383 (0.141)***
Military Action		
UN Endorsement*Military Action		
Constant	6.702 (0.081)***	7.366 (0.077)***
N of Observations	4374	4374
N of Subjects	1108	1108
	Model 8	Model 9
UN Endorsement Treatment	0.773 (0.099)***	0.572 (0.133)***
Country Named		-0.069 (0.093)
UN Endorsement*Country Named		-0.003 (0.133)
Security Threat		-1.409 (0.090)***
UN Endorsement*Security Treat		0.398 (0.133)**
Military Action	-1.452 (0.095)***	-1.456 (0.090)***
UN Endorsement*Military Action	0.166 (0.140)	0.176 (0.133)
Constant	7.389 (0.077)***	8.131 (0.099)***
N of Observations	4374	4374
N of Subjects	1108	1108

Dependent variable: support score for intervention measure (0=strongly oppose, 10=strongly favor)

One way to unpack the functional form problem is to study subsets of subject whose baseline support for humanitarian interventions and security interventions are the same. I identified three such groups: 1) Republicans, 2) subjects who think the UN does a poor job, and 3) subjects who have little confidence in the Executive Branch. (See Table 11) After performing t-tests, I determined the baseline support for humanitarian crisis interventions and the baseline for security threat interventions are not significantly different in each group. Next, I performed the interaction term regression analysis from Model 7 on each of the three groups. The results are reported in Table 12.

Table 11: Three groups with similar baseline support

	1. Republicans	2. Subjects who think the UN does a poor job	3. Subjects who have little confidence in the Executive Branch
Humanitarian Crisis Baseline	7.538	7.096	6.814
Security Threat Baseline	7.534	6.658	6.229
Difference	0.004	0.438	0.585
Standard Error	0.326	0.541	0.388
t-statistic	0.012	0.809	1.506
Degrees of Freedom	235	147	234

Table 12: GLS regression: random effects; three groups with similar baseline support

	Model 10: Republicans	Model 11: Subjects who think the UN does a poor job	Model 12: Subjects who have little confidence in Executive Branch
UN Endorsement Treatment	0.340 (0.262)	0.157 (0.422)	0.721 (0.018)**
Security Threat	0.110 (0.258)	-0.442 (0.410)	-0.566 (0.294)*
UN Endorsement*Security Treat	0.323 (0.374)	0.428 (0.593)	0.135 (0.432)
Constant	7.503 (0.224)***	7.140 (0.368)***	6.790 (0.273)***
N of Observations	487	306	476
N of Subjects	122	77	121

Dependent variable: support score for intervention measure (0=strongly oppose, 10=strongly favor)

As Models 10 through 12 shows, the interaction effect of UN endorsement and security threat is not significant in any of the three groups. This suggests my initial findings of a greater ATE for security interventions might be attributed to the functional form rather than the intervention type. While the three groups in Table 12 are not entirely representative of the entire sample, they are different enough – both politically and in their baseline support – to suggest the functional form problem exists for other subjects as well.

The functional form problem poses a great challenge for experimental IR scholars in general and should not be taken lightly. The main reason is that the public tend to give greater baseline support for humanitarian interventions than for security interventions (Jentleson & Britton 1998). A similar argument can be applied to those who are studying whether UN endorsement matters more in low versus high-cost interventions. Low-cost interventions will probably have a higher baseline than high-cost interventions. One possible way to overcome the functional form problem is to assign a fixed baseline to all subjects. For instance, all subjects are told that without UN endorsement, their support for each intervention measure is a 5. Then for each randomly assigned intervention measure, subjects are asked what their support score would be given UN endorsement. If researchers intend to study only the variation of ATE by intervention type, they should submit all their subjects to the UN endorsement treatment, use a fixed baseline, and vary the type of intervention measure.

3.4.3 *Who listens to the UN?*

In general, UN endorsement raises subjects' support for intervention measures. But it is probably true that some subjects experienced greater ATE than others. The results shown in Table 13 and Table 14 suggest UN endorsement matters more to subjects who support the UN. To test for heterogeneous treatment effects, I first calculated the CATEs for each subset of subjects using r_i . Next, I tested whether those differences in CATEs were significant using feasible GLS regressions with interaction terms.

Table 13: Conditional average treatment effects (CATEs)

	Non-Isolationists	Strong Isolationists	Too Much Spending	Too Little Spending
Military Action ATE	1.05	1.58	1.25	1.67
SE	0.41	0.69	0.14	0.65
N of observations	171	110	1551	68
Economic Sanction ATE	0.35	1.69	0.80	0.09
SE	0.35	0.67	0.13	0.49
N of observations	172	110	1551	68
	Low Support for the UN	Strong Support for the UN	Strong Confidence in Executive	Weak Confidence in Executive
Military Action ATE	0.46	1.37	1.29	1.23
Std. Dev.	0.89	0.23	0.22	0.39
N of observations	48	284	532	267
Economic Sanction ATE	0.37	1.35	0.81	0.27
SE	1.00	0.22	0.19	0.36
N of observations	48	284	532	267
	Republicans	Democrats		
Military Action ATE	0.72	1.07		
Std. Dev.	0.32	0.14		
N of observations	153	1388		
Economic Sanction ATE	0.18	0.85		
Std. Dev.	0.29	0.12		
N of observations	154	1388		

Dependent variable: support score for intervention measure (0=strongly oppose, 10=strongly favor)

Table 14: GLS regression: random effects; heterogeneous treatment effects

	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
UN Endorsement Treatment	0.836 (0.149)***	0.728 (0.277)**	0.897 (0.079)***	-0.109 (0.238)	0.640 (0.166)***	-0.391 (0.463)
Isolation	-1.828 (0.243)***					-1.654 (0.274)***
UN Endorsement*Isolation	0.013 (0.299)					0.297 (0.342)
Government Spending		-1.303 (0.254)***				-0.633 (0.281)*
UN Endorsement*Government Spending		0.189 (0.310)				-0.012 (0.352)
Republican			0.996 (0.197)***			0.626 (0.243)*
UN Endorsement*Republican			-0.385 (0.235)			0.042 (0.301)
Support for UN				-0.472 (0.282)		-0.541 (0.310)
UN Endorstmnt*Support for UN				1.471 (0.337)***		1.692 (0.384)***
Confidence in Executive					0.359 (0.218)	0.629 (0.250)*
UN Endorsement*Confidence in Executive					0.407 (0.137)	0.058 (0.313)
Constant	7.465 (0.121)***	7.763 (0.226)***	6.557 (0.065)***	7.000 (0.199)***	6.486 (0.137)***	7.865 (0.371)***
N of observations	4258	3719	4374	4094	4159	3411
N of groups	1078	941	1108	1035	1053	861

Dependent variable: support score for intervention measure (0=strongly oppose, 10=strongly favor)

Table 13 tells a complex story. One of Grieco et al.'s (2011) conclusions does not hold up. Subjects who have little confidence in the Executive Branch actually experienced smaller ATE than those who have stronger confidence. Similarly Model 17 (Table 14) demonstrates that the interaction term of UN endorsement and confidence in the Executive Branch is not significant. This confirms my suspicion that Grieco et al.'s survey experiment might be influenced by the political events of the time. At the time my survey experiment was conducted, Barack Obama served as the president of the U.S. Those who have strong confidence in him are likely to be Democrats, who are generally more supportive of the UN than Republicans (Kohut & Stokes 2007). But in 2003, the people who demonstrated strong confidence in the Executive Branch are most likely Republicans. Thus the story Grieco et al. tells might have less to do with distrust in the president than with people's partisanship.

The CATE of UN endorsement is higher for Democrats than for Republicans, as Table 14 shows. The CATE in cases of economic sanction for Democrats is more than 4 times as large as those for Republicans, although the differences between the CATEs in cases of military action is not so great. As Model 15 (Table 14) shows the difference between the Republican and non-Republican CATEs is not statistically significant. This occurred because I pooled the military action and economic sanction observations together for the regression analysis in Table 14. When I ran the regression with only economic sanctions, there exists a significant difference between the CATEs of Republicans and non-Republicans.

As I suggested earlier, party labels might simply be proxies for subjects' attitudes towards the UN. To test this, I examined the CATEs for those who strongly support the UN versus those who have low support for the UN.⁵ The CATE of subjects who strongly supported the UN is almost 3 times as large as the CATE of those who expressed low support for the UN. Similarly, the CATE of strong UN supporters is 5 times as large compared to that of weak UN supporters. Model 16 demonstrates that the difference between the two groups' CATEs is statistically significant. It is far from surprising that the strong UN

⁵ This variable is generated based responses to a question that asks how much power should the UN have. I classify subjects who want the UN to have great amounts of power as "strong UN supporters" and those who want the UN to have little power as "weak UN supporters."

supporters are the ones who are most affected by UN endorsements. According to Lupia and McCubbins (1998), elite cues only work if the cue giver and receiver share common interests. The subjects who responded the most to the UN cue might have done so because they share preferences with the UN.

The CATEs of non-isolationists and strong isolationists differ little overall. Non-isolationists, most likely, always want to support intervention; at the same time, they are more likely to be internationalist in outlook. In contrast, isolationists are more likely to reject interventions in general; however, they might be less happy with U.S. engagement with IOs. These complex factors might have canceled out the differences in CATEs between the two groups.

Similarly, the CATEs of subjects who think the government spends too much and those who think the government spends too little are not statistically different either. Initially, I hypothesized that subjects who are very cost averse are more likely to be affected by UN endorsement because they have more movement between control and treatment. UNSC sanction for war signals merit, and therefore mitigate these subjects' fears of wasting money on an unwarranted war. Conventional IR theory asserts that UNSC sanctions' power come from their ability to signal merit (i.e., the military mission is worth its cost) (Thompson 2006). But Tingley and Tomz's study have discounted that argument. After their treatment/control prompts, the authors asked their subjects a series of mediation questions about the material costs and benefits of sending military forces to measure perceptions about the merit of going to war. If the public considers UN resolutions as signals of merit, their perceptions of merit would vary depending on the treatment assignment; however, the authors do not find such a variation. This finding can serve as one reason that I do not observe a difference in the CATEs of subjects whose opinions about government spending vary.

In sum, studying the CATEs of different groups raises some interesting questions about the nature of UN cues. The significant difference in CATEs between those who strongly support the UN versus those who do not suggests UN cues trigger something other than

rationalist thinking in the minds of Americans. If UNSC sanctions don't signal merit, then what do they do? Tingley and Tomz (2012) suggest that it creates a public commitment that the U.S. must fulfill. This theory is certainly plausible and has evidence in experimental data. As an extension to this reasoning, I argue that Americans who value international organizations feel more strongly about the U.S.'s obligation to fulfill public commitment. These Americans are probably more cosmopolitan in their outlook and more responsive to non-American elite cues, like those liberal Democrats Hayes and Guardino (2011) described.

3.4.4 *External validity check*

Finally, I checked the external validity of my survey experiment by testing to see if adding country names to the scenario prompts causes UN approval to have a different treatment effect. The two country names included in the security threat scenarios are North Korea and Iran. The treatment country names in the humanitarian crisis scenarios included Chad, Angola, Sudan, South Sudan, Burma, Burundi, Laos, Uganda, the Democratic Republic of the Congo, and the Central African Republic. I performed a difference-in-difference analysis comparing UN approval's treatment effect in generic prompts and specific prompts. The results are presented in Model 6 (Table 10). The interaction term of UN endorsement and country name appearing in the prompt is statistically insignificant. This means adding country names to scenario prompts does not affect the treatment effect of UN approval.

The results are rather promising for scholars who wish to test theories of international relations. Often, researchers feel conflicted about whether to present generic prompts (e.g., "Country A plans to go to war") or specific prompts (e.g., "Pakistan plans to go to war"). A randomly rotating list of country names, even as small as two countries, in specific prompts can produce results similar to those of a generic prompt. One reason is that, on average, the specific country effects cancel each other out when they are put into a randomly rotating list. Another reason is that subjects are already thinking about specific

countries when they read a generic prompt. To test the second hypothesis, one can assign subjects a generic prompt then ask them which country names came to their mind. Nevertheless, I offer caution to those who want to engage in this practice because one must choose realistic country names. For instance, it would be imprudent to write that Sweden threatens the U.S. with a nuclear bomb. Unrealistic scenarios are likely to confuse and distract subjects, especially those who are more knowledgeable about world affairs.

3.5 CONCLUSION

The results from the Yale pilot study largely confirm my hypotheses about how UN approval can influence public opinion concerning military interventions and economic sanctions. Through this survey experiment, I arrive at four conclusions. First, UN approval for economic sanctions increases public support for them. Second, studying the ATEs of UN endorsements across different types of interventions is difficult because the ATEs of UN cues (or other elite cues) follow a logistic distribution. Future scholarship focused on this specific question should employ a fixed support baseline. Third, I discover that UN endorsements have a larger CATE on people who strongly support the UN; CATEs vary little among other subsets of the subject. Furthermore, I reject Grieco et al.'s theory that people with low confidence in the president are the ones most moved by UN cues. Four, I conclude that introducing country names to scenario prompts using a randomly rotating list has no effect on the ATE of UN approval.

This pilot study is not ideal for several reasons. First, Yale students are not representative of the U.S. public. Second, I relied on a convenience sample. Through Mechanical Turk, I hope to get a sample that is more representative of the U.S. population. I acknowledge there are flaws associated with that web service as well. But recent research suggests that Mechanical Turk respondents, though less representative of the U.S. population than commercial survey panels, are often more representative than in-person convenience samples (Berinsky, Huber & Lenz 2012).

In the next stage of my research, I plan to study why some Americans and not other pay attention to international elites. Hayes and Guardino's (2011) work, based on observational data, suggests that certain subsets of Americans – politically informed Democrats and Independents – are more predisposed to the influence of foreign media, when compared with the general public. Through experimental methods, I hope to unpack the reasons why these subgroups might experience larger CATE when exposed to non-American elite cues. If IO cues signal public commitment rather than merit, as Tingley and Tomz argue, American public opinion of foreign policy might be influenced more by the desire to maintain reputation than by rationalist calculations about material costs/benefits. I hypothesize that cosmopolitan Americans are more likely to be influenced by foreign elites because they care about the U.S.'s reputation as a country committed to international cooperation.

3.6 ACKNOWLEDGMENTS

I thank Alan Gerber, Allan Dafoe, Nikolay Marinov, Peter Aronow, John Bullock, Vinicius Lindoso, Dustin Tingley, and Michael Tomz for their helpful comments and suggestions.

4 FOREIGN OPPOSITION TO WAR AND THE AMERICAN PUBLIC

Abstract. Scholars have observed that foreign opposition to war depresses domestic support for the conflict. By the logic of Crawford-Sobel's model of strategic communication, opposition by allies should be more informative and credible than opposition by rivals. Thus, I hypothesize a UN Security Council (UNSC) veto by an American ally (France) will decrease support for war more than a veto by an American rival (Russia). Furthermore, I hypothesize Democrats will express greater objection to a war than Republicans when there exists foreign opposition to the war. In an online survey experiment, I test whether variations in UNSC vetoes affect subjects' attitudes about a hypothetical war proposed by the U.S. The experimental data demonstrate that the French veto and the Russian veto lowered public support for war by the same amount. In addition, I find that Democrats, more than Republicans, were more persuaded by foreign opposition to the war.

4.1 INTRODUCTION

The invasion of Iraq in March 2003 by U.S.-led coalition forces was not without international controversy. Russia, China, and even U.S. allies, France and Germany, opposed the war during its planning stages.¹ France made it clear it would use its United National Security Council (UNSC) veto against a UN resolution that gave Iraqi leader Saddam Hussein until March 17, 2003 to disarm. This resolution contained an ultimatum that would lead to war.²

Americans responded to foreign opposition to the Iraq War in divergent ways. On one hand, Republican elites launched a wave of anti-French messages; among the most memorable was the Republican-controlled Congress's decision to rename French fries "Freedom fries" in Congressional cafeterias. Republican representative Bob Ney said the action was "a small but symbolic effort to show the strong displeasure of many on Capitol Hill with the actions of our so-called ally, France."³ This elite sentiment was reflected in public opinion towards France. According to the QUS10b Pew Global Attitudes survey, in 2003, 60 percent of Americans viewed France unfavorably and only 29 percent viewed France favorably.⁴

On the other hand, a subsection of the American public was influenced by foreign elites to oppose the Iraq War. Hayes and Guardino's (2011) analysis of public opinion surveys from August 2002 through March 2003, shows that Democrats and Independents with high levels of political awareness were swayed by foreign elites opposing U.S. proposal for war. Furthermore, some Americans valued approval by the UNSC. In a USA TODAY/CNN/Gallup poll conducted three days before the invasion, 58 percent of Americans favored invading Iraq to remove Saddam Hussein from power. But the support dropped to 47 percent if the Bush administration did not seek a final UNSC vote.⁵

1 January 23, 2003. "Opposition to Iraq war." *BBC News*.

2 March 10, 2003. "France will use Iraq veto." *BBC News*.

3 March 12, 2003. "US Congress opts for 'freedom fries.'" *BBC News*.

4 "America's Place in the World 2009: An Investigation of Public and Leadership Opinion About International Affairs." December 2009. Pew Research Center for the People and the Press.

5 Benedetto, Richard. March 16, 2003. "Poll: Most back war, but want U.N. support." *USA Today*.

The interaction between foreign elites and the American public in the lead-up to the Iraq War poses an interesting question. Do foreign elites influence domestic politics? Traditionally, scholars had ignored the role of international elites in the intersection between international relations and domestic politics; instead, they focused on domestic leaders and diplomats. Elected leaders worry about domestic audience cost when deciding to take military action against another state (Fearon 1994). Diplomats in international negotiations engage in a two-level game with other diplomats and with their domestic government (Putnam 1988).

But a more recent line of research centers on the indirect impact of IOs on domestic public opinion. In democratic countries, the attitudes of the mass public constrain leaders to behave in certain ways. For instance, when a president or prime minister tries to persuade the public to go to war, the public can get a “second opinion” from the United Nations Security Council. In effect, authorization by the UN Security Council legitimizes military interventions in the eyes of the masses. Chapman and Reiter’s (2004) study of historical polling data and Grieco et al. (2011) and Tingley and Tomz’s (2012) survey experiments contend that the UN affect mass public support for war.

I contribute to this small but growing literature in two ways using a survey experiment. For one, I expand upon previous studies to consider how individual countries affect public support for war. There are many ways a U.S.-backed UNSC resolution could fail to pass. Opposition could come from an American ally, an American rival, or a combination of both. I hypothesize that variations in opposition would induce different levels of support for war. Following the logic of Crawford and Sobel’s (1982) model of strategic communication, opposition from an ally would be more informative about the merit of the proposed war than opposition from a rival. Therefore, the American public would express less support for a military campaign when an ally opposes than when a rival opposes.

Nevertheless, one cannot presume that the American public will behave in a fully rational way. Therefore, I study whether foreign opposition to a U.S.-proposed war would have heterogeneous treatment effects on public opinion. Frequently, Americans view foreign

policy and international news through a partisan lens (Berinsky 2009). Because Republicans tend to be more nationalistic and less cosmopolitan in their outlook than Democrats, I predict foreign opposition would be more persuasive to Democrats than Republicans.

Interestingly, experimental data from the survey experiments contradict some of my hypotheses. The French veto and the Russian veto lowered public support for war by the same amount. But true to my second hypothesis, I find that Democrats, more than Republicans, were more persuaded by foreign opposition to the war.

4.2 LITERATURE REVIEW

4.2.1 *The influence of foreign elites*

Most of the IR research about the influence of foreign elites on domestic publics have concerned sanctions or resolutions by IOs. There exists much theoretical work about why states consult formal IOs before engaging in international conflicts. It is widely assumed that because IOs hold a degree of independence from other states, they provide the most unbiased information. Therefore, by obtaining approval from the UN Security Council, a state can cast its military actions — even unilateral ones — as legitimate (Abbott & Snidal 1998). Voeten (2005) expands upon this theory by arguing the Security Council's authorizations serve as an information shortcut for the domestic audience. Americans generally want their country to be involved internationally, but they also fear overextension of the U.S. military. An authorization by the UN indicates that no costly challenges will result from the military action.

This theory of how IOs influence public opinion about military interventions has support in empirical evidence from a study of historical data and two survey experiments. Chapman and Reiter (2004) examine polling data to see how Americans responded to involvements in militarized interstate disputes between 1945 and 2001. They note that presidential approval ratings were 9 percentage points higher, on average, in missions

that had UN Security Council approval. Although Chapman and Rieter find a positive correlation between UN approval and public support for the president (implicitly the military intervention), one cannot conclude that the former caused the latter. For instance, reverse causality is possible: historically, the Security Council might have simply approved the interventions that everyone, including the mass public, agrees is legitimate. Grieco et al. (2011) and Tingley and Tomz's (2012) survey experiments overcome the limitations of Chapman and Reiter's observational study.

The data for Grieco et al.'s (2011) paper came from a 2003 survey experiment conducted over the phone. The subjects were presented with a prompt that described the hypothetical invasion of East Timor by Indonesia. The treatment group received a cue indicating the UN and NATO supported a U.S. military mission in East Timor. In contrast, the control group received a cue that stated the UN and NATO does not approve of such an operation. True to the authors' hypothesis, the IO cue has a significant effect on increasing individual subjects' support for military intervention in East Timor. The effect is most notable among those who valued the UN and NATO and those who had little confidence in the president. Grieco et al. admit the shortcoming of their survey experiment's external validity: the prompt they presented is representative only of low-cost humanitarian interventions. Tingley and Tomz's (2012) study furthers the discussion by testing why UN Security Council authorization affects public opinion. They hypothesize three reasons: first, the UN cue suggests military force is warranted; second, the UN cue indicates that other countries will share the military burden; third, the UN cue signals a public commitment to contribute to a multinational effort. The third reason is the most consistent with the data from the authors' survey experiment.

Following this line of work, I use variations in UNSC votes as treatments in my survey experiment. In general, it is difficult to create a treatment that operationalizes the construct of country opposing a U.S.-backed war. Suppose one creates a vignette that states, "France opposes the U.S. proposal to invade Country X." Survey subjects might be confused about who or what is opposing the war: French voters, the French government, the president of

France? Therefore, UNSC vetoes serve as good, formal proxies for countries' opposition to a proposed military campaign. These statements are unambiguous and easy for readers to interpret.

4.2.2 *Informative versus non-informative cheap talk*

One of the most ironic aspects of UNSC votes is the disconnect between diplomats' statements about the resolutions they pass and the reality on the ground. Delegates to the UN say they broadcast credible signals when their resolutions are merely cheap talk.

Diplomats often use the language of signaling when referring to resolutions. For instance, when the UNSC unanimously adopted Resolution 1441 to offer Saddam Hussein a final opportunity to disarm, Secretary-General Kofi Annan said, "Whenever the Council is united, it sends a very powerful signal and I hope that Iraq will heed that signal."⁶ The term "signal" that the diplomats want to convey is analogous to the term used in contract theory. Signaling, as defined in Spence (1973), is the transfer of information from one party to another. Within a signaling game, one player is informed and the other is not. The informed player's strategy set consists of signal contingent on information and the uninformed player's strategy set consists of actions contingent on signals. In Spence's job market example, there exists a signaling cost for workers: those who want jobs obtain education to credibly signal to employers.

But are UNSC votes costly signals? They do not create international law and only occasionally create binding legal obligations. They are prime examples of the signals that Jervis (1989) discusses: "They do not contain inherent credibility. They do not, in the absence of some sort of enforcement system, provide their own evidence that actors will live up to them" (18). In some ways, they can almost be considered cheap talk because the communication is costless. In a cheap talk game, a special variant to the signaling game, communication do not directly affect payoff.

⁶ November 8, 2002. "Secretary-General's statement at the adoption of Security Council Resolution 1441 on Iraq."

Suppose there are two players, a sender and a receiver. If the two have opposing preferences, the only equilibria are “babbling” equilibria. Any information the sender sends is uninformative and ignored by the receiver. Yet, cheap talk can be informative when players’ preferences are close (Crawford & Sobel 1982, Crawford 1998). This is especially true if the sender sends a message that goes against his own self-interest. If the sender and the receiver have identical preferences, the cheap talk game would turn into a coordination game. In the equilibrium, the sender would send truthful information and the receiver would believe him or her.

If the logic of Crawford-Sobel’s strategic communication model exists in international relations, then some UNSC votes are more informative than others. For example, the U.S. and France have similar preferences in foreign policy. In contrast, American and Russian foreign policy interests are not as similar. According to the Crawford-Sobel model, a veto by France would be more informative than a veto by Russia. As a result, the French veto might become a more salient issue among the American public and have a greater effect on their support for war. Nevertheless, one should not expect the American public to react in a fully rational fashion given previous findings about their inconsistencies towards policy issues (Zaller 1992, Converse 2006). Therefore, in my survey experiment, I test whether the Crawford-Sobel model holds up in a public opinion context.⁷

4.3 HYPOTHESES AND RESEARCH DESIGN

4.3.1 *Hypotheses*

My survey experiment seeks to answer two questions. First, do variations in UNSC votes affect public support for a U.S.-backed proposal for war? Second, does foreign opposition to a U.S.-backed proposal for war have heterogeneous effects on American voters’ attitudes?

⁷ I acknowledge Allan Dafoe for pointing out to me the connection between the variation in UNSC votes and the Crawford-Sobel model.

Each treatment in my survey experiment consists of a vignette describing a U.S.-backed proposal for war against an aggressor country that is invading its neighbor. There are four experimental groups. Subjects in all four groups read about a UNSC vote on a resolution to take military action against the aggressor country. In all cases, the U.S., U.K., and China vote in the affirmative. I vary the French and Russian votes:

	Russia votes YES	Russia votes NO
France votes YES	Control group	Treatment 1
France votes No	Treatment 2	Treatment 3

I denote $Y(F, R)$ as the subjects' level of support for the U.S.-backed proposal for war. Let F be France, R be Russia, Y for a "yes" vote and N for a "no" vote. I hypothesize the following:

Hypothesis 1: According to the logic of Crawford-Sobel's strategic communications model, $Y(Y, Y) > Y(Y, N) > Y(N, Y) > Y(N, N)$.

Subjects' support for war is predicted to be highest when both France and Russia vote "yes" $Y(Y, Y)$. This is due the fact that UNSC resolution passes since none of the permanent member countries vetoes it. As Grieco et al.'s (2011) and Tingley and Tomz (2012) have demonstrated in their experiments, public support for wars are higher when the UN endorses them. Subjects' support for war is predicted to be higher when Russia vetoes $Y(Y, N)$ than when France vetoes $Y(N, Y)$. I hypothesize this phenomenon exists because I think subjects will consider opposition by an American ally to be a more credible message than opposition by an American rival. In subjects' minds, the French veto might mean the proposed war lacks merit or that other American allies might not aid in the military campaign. In contrast, the Russia veto is rather like cheap talk because Russia is a historical U.S. rival and has vetoed/abstained on numerous U.S.-backed UNSC resolutions. Therefore, it would be difficult to draw credible information from the Russian veto. Finally, I

predict that the combined vetoes by France and Russia $Y(N, N)$ will cause support for war to drop to the lowest level.⁸

Aware of the Freedom Fries phenomenon, I predict my treatments will induce heterogeneous treatment effects among the subjects. I predict that subjects of different party identification will have different reactions to the treatments.

Hypothesis 2: The conditional average treatment effect (CATE) of opposition to the proposed war will be higher for Democrats than Republicans.

I define the average treatment effects to be the drop in support for the proposed war when the UNSC vetoes a resolution. Therefore, the ATEs are $E[Y(Y, Y) - Y(Y, N)]$, $E[Y(Y, Y) - Y(N, Y)]$, and $E[Y(Y, Y) - Y(N, N)]$. Formally, my second hypothesis states that

- $E[Y(Y, Y) - Y(Y, N)|\text{Democrat}] > E[Y(Y, Y) - Y(Y, N)|\text{Republican}]$
- $E[Y(Y, Y) - Y(N, Y)|\text{Democrat}] > E[Y(Y, Y) - Y(N, Y)|\text{Republican}]$
- $E[Y(Y, Y) - Y(N, N)|\text{Democrat}] > E[Y(Y, Y) - Y(N, N)|\text{Republican}]$.

Democrats and Independents are more sensitive to foreign opinions than Republicans. Studies have shown that Democrats and Independents are more likely to endorse collaborative decision making between the U.S. and the European Union and are more willing to cede U.S. autonomy in foreign policy decisions than Republicans (Holsti 2004, Page & Bouton 2006). Furthermore, Democrats and Independents were influenced by foreign elites opposing the Iraq War while Republicans were not influenced (Hayes & Guardino 2011). Therefore, I predict that Democrats and Independents will be persuaded by the foreign opposition to a greater degree than Republicans.

4.3.2 *Survey experiment design*

To shed new light on how variations of UNSC votes affect American public opinion, I design a new experiment and embed in a public opinion survey. I field this survey on

⁸ I acknowledge Allan Dafoe for writing out this formal notation.

Amazon.com's Mechanical Turk, an online crowd-sourcing service. I acknowledge there are flaws associated with that web service. Recent research suggests that Mechanical Turk respondents, though less representative of the U.S. population than commercial survey panels, are often more representative than in-person convenience samples (Berinsky, Huber & Lenz 2012). I recruit a sample of 800 subjects, with 200 subjects in each experimental group.

The survey begins with some basic demographics questions taken from the American National Elections Studies (ANES). Subjects are asked about their age, sex, political ideology, level of education, foreign language skills, and party identification. After the demographic questions, subjects are asked two questions to determine whether their beliefs about foreign countries match reality. In short, I want to check whether subjects think France is a U.S. ally and Russia is not. Subjects are asked to identify U.S. military allies from a list of 10 countries. Next, subjects are asked to identify countries they think threaten U.S. national security interest from the same list.⁹

Afterwards, subjects are asked to read a short passage about the UNSC. They are told they will be quizzed on the information later. I tell subjects beforehand they will receive an extra 25 cents in pay if they get all three quiz questions correct. The passage about the UNSC reads:

The U.N. Security Council has 5 permanent members, who have been on the Council since 1946. The 5 permanent members are the United States, the United Kingdom, France, China, and Russia.

The Council also has 10 temporary members, who are elected every two years by the U.N. General Assembly. This year, the temporary members are Azerbaijan, Argentina, Australia, Guatemala, South Korea, Morocco, Pakistan, Luxembourg, Rwanda, and Togo.

The Council sometimes passes resolutions about the use of military force. When deciding whether to pass a resolution, each of the 15 members gets one vote. A resolution passes if at least 9 members vote YES. However, if any permanent member votes NO, the resolution fails. Thus, each of the 5 permanent members has the power to block or "veto" a resolution.¹⁰

⁹ The order of the countries in the list is randomized. The countries include Japan, United Kingdom, Russia, Egypt, Brazil, Pakistan, Iran, China, France, and Venezuela.

¹⁰ This passage about the UNSC was adapted from Tingley and Tomz's (2012) study.

When subjects advance to the next web page to take the quiz, they cannot go back to the passage they have just read. The three multiple-choice questions I ask subjects are 1) Which 5 countries are permanent members of the UNSC? 2) What happens if 1 of the 5 permanent members votes NO on a resolution? and 3) How many countries are temporary members of the UNSC?

After completing the quiz, subjects advance to the experimental section of the survey. They read the following instructions: “We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.”

All subjects are then presented with an international crisis: “A country recently sent its military to take over a neighboring country. The attacking country is led by a dictator, who invaded to get more power and resources. The attacking country has a powerful military, and the neighbor is too weak to defend itself. The U.S. president says the invasion is immoral and will hurt U.S. interests. He wants the U.S. military to push out the invaders.”¹¹ The scenario was loosely based on previous UNSC resolutions involving the use of force, such as measures to stop the Iraqi invasion of Kuwait in 1990.

Next, subjects were told about how members of the UNSC voted on a resolution for military action against the aggressor country. The four experimental conditions are:

1. Control: All 5 permanent members vote YES and the resolution passes.
2. Treatment 1: The resolution fails because Russia votes NO.
3. Treatment 2: The resolution fails because France votes NO.
4. Treatment 3: The resolution fails because Russia and France vote NO.

Here is an example of one vignette (Treatment 3) a subject would read:

The United States proposes a resolution to the United Nations Security Council (UNSC) to use military force against the aggressor country. The UNSC fails to pass the resolution because two permanent members of the UNSC vote NO on it.

¹¹ This vignette has previously been used in Tingley and Tomz’s (2012) study.

Here are the votes of the five permanent members of the UNSC.

YES on the resolution: United States, United Kingdom, China

NO on the resolution: France, Russia

After reading the vignettes, subjects are asked whether they agree or disagree that the U.S. should use military force to push out the invaders. Subjects respond on a five-point scale: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. Subjects can also select “I don’t know” as an option. Afterwards, subjects are asked to write why they selected the choice they selected.

Finally, subjects answer additional demographics questions, including their ethnicity and how much they consume international news. I assume these questions will not be affected by the treatments. I did not want to front-load the survey with too many demographics questions because I fear subjects’ mental capacity to process the treatment vignettes would be impaired.

4.4 RESULTS AND DISCUSSION

The data from my survey experiment fail to confirm my first hypothesis but confirm parts of my second hypothesis. Subjects who received the treatment of the French veto expressed similar levels of support for the proposed war as subjects who received the treatment of the Russian veto. Nevertheless, the veto by both Russia and French depressed public support for war the most. I also observe heterogeneous treatment effects among difference partisan subjects. Democrats, more than Republicans, experienced larger drops in public support for the war in each treatment.

Before proceeding to the analysis of this survey experiment, I describe the data I have collected. Next, I discuss the implications of my data for the two hypotheses previously stated.¹²

¹² My pre-analysis plan is available upon request.

4.4.1 Data

The survey experiment was conducted in early April 2013. Using Amazon.com's Mechanical Turk, I obtain 800 subjects, of which 790 completed the survey. For the purpose of this analysis, I only consider subjects located in the U.S. and are American citizens. Given these conditions, I derive a final sample of 784 subjects. Two hundred subjects are in the control group; 205 are administered Treatment 1 (the Russian veto), 184 are administered Treatment 2 (the French veto), and 195 are administered Treatment 3 (the Russian and French vetoes).

The dependent variable of interest is each subject's support for the U.S. proposed war. I label this variable *support score*. The *support score* is derived from the question following the treatment vignettes. In the question, subjects are asked how much they agree or disagree that the U.S. should use military force to push out the invaders. I construct the *support score* in this manner:

- Strongly disagree: *support score*=1
- Somewhat disagree: *support score*=2
- Neither agree nor disagree: *support score*=3
- Somewhat agree: *support score*=4
- Strongly agree: *support score*=5

Twenty-two subjects selected "I don't know" as answers to the question. Rather than imputing their *support score*, I will not use these subjects' responses in my analysis. I conduct a Person's chi-squared test to determine if the "I don't know" responses are independently distributed across the four experimental conditions. The χ^2 statistic is 1.7848, with a p-value of 0.618. This suggests the "I don't know" responses are independently distributed among the control and the treatment conditions; therefore, I can drop these responses without too much concern.

The pre-treatment covariates I include for analysis are:

- age: continuous variable (minimum=18, maximum=73)
- sex: indicator variable (0=female, 1=male)
- education: indicator variable (0=no college degree, 1=college degree)
- party identification: indicator variable (0=Republican, 1=Democrat)¹³
- political ideology: 7-point scale (1=very liberal, 7=very conservative)

To check the random assignment procedure is sound, I check for balance among the covariates¹⁴ As Table 15 shows, the pre-treatment covariates are essentially the same across the four experimental groups. Furthermore, when I regress the treatment assignment on these covariates, the beta coefficient for each is not significant. Thus, I conclude that my randomization procedure has worked and no administrative errors were committed.

Table 15: Covariates balance check

Pre-treatment covariates	Control: Both YES	T1: Russia NO	T2: France NO	T3: Both NO
Mean age	32	31	31	32
Proportion college graduates	0.51	0.45	0.51	0.47
Proportion Democrats	0.64	0.56	0.61	0.63
Mean ideology score	3.25	3.40	3.21	3.30
Proportion females	0.54	0.56	0.53	0.56

4.4.2 *Effect of foreign opposition*

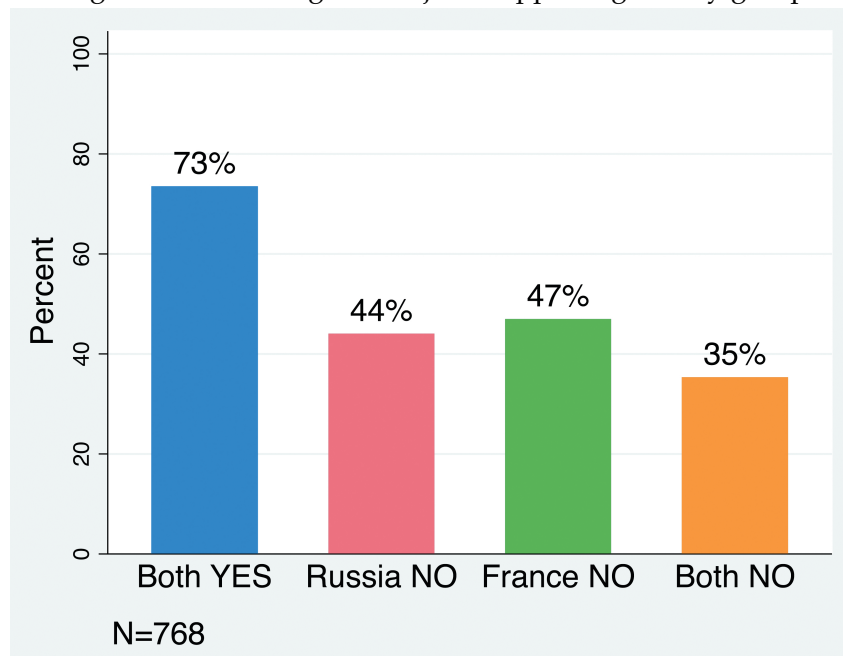
In the first part of my analysis, I test whether the variations in experimental conditions induce different levels of support for the U.S. proposed war. In Figure 11, I compare the

¹³ I define Republican as subjects who label themselves as “strong Republican,” “weak Republican,” and “independent, leaning Republican.” Likewise, I define Democrats as subjects who label themselves as “strong Democrat,” “weak Democrat,” and “independent, leaning Democrat.”

¹⁴ The randomization is done using Qualtrics’s online software built-in randomization tool.

percentage of subjects supporting the war in each experimental group.¹⁵ In the treatment condition, when both Russia and France support the war, public support is rather high. The percentage of subjects who support war drops significantly in each of the three treatment conditions. Contrary to my hypothesis, the percentage of those who support war is slightly higher in the French veto group than in the Russian veto group. The percent who support war is lowest in the Russian-and-French vetoes group.

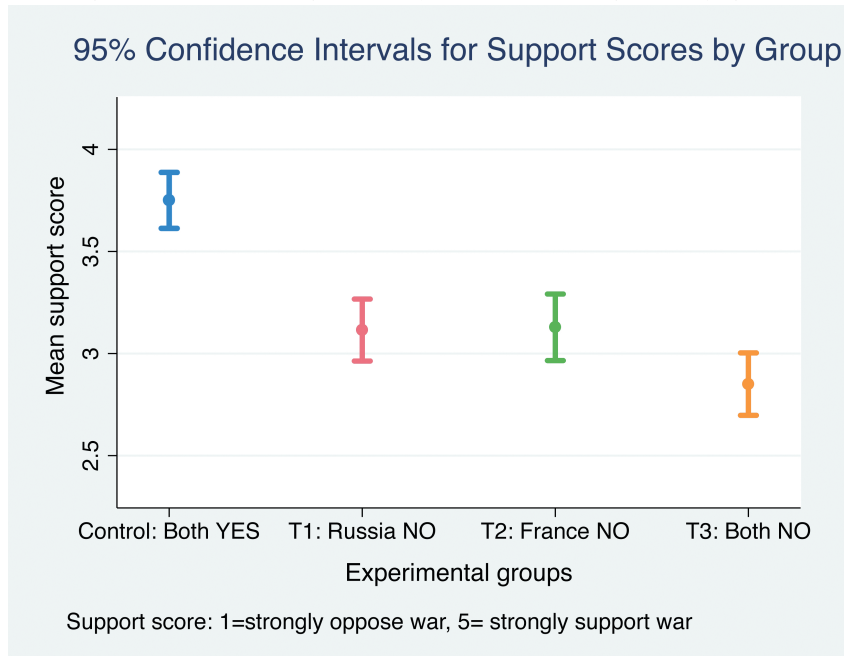
Figure 11: Percentage of subjects supporting war by group



In Figure 12, I compare the *support scores* of the four experimental groups. The figure contains the mean *support score* and the 95 percent confidence interval for each group. Once again, the outcome variable appears to be very similar between the Russian veto group and the French veto group. If the *support scores* are statistically the same in these two groups, then my first hypothesis is incorrect.

¹⁵ I define supporting the war as having a *support score* of 4 or 5.

Figure 12: Percentage of subjects supporting war by group



To formally test Hypothesis 1, I set up the following linear model:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + u \quad (13)$$

where b_0 is the constant, X_{1i} is a dummy variable that equals 1 when Russia vetoes, X_{2i} is a dummy variable that equals 1 when France vetoes, X_{3i} is a dummy variable that equals 1 when both Russia and France vetoes (essentially an interaction term), and X_{4i} represents the pre-treatment covariates include.¹⁶ To test my hypothesis that $Y(Y, Y) > Y(Y, N) > Y(N, Y) > Y(N, N)$, I conduct one-way t-tests on the coefficients. If my hypothesis is correct, $b_0 > b_0 + b_1 > b_0 + b_2 > b_0 + b_1 + b_2 + b_3$. The results without pre-treatment covariates adjustment are reported in Table 16. The results with pre-treatment covariates are reported in Table 17.

Without covariates adjustment, the drop in *support score* is 0.622 points for the French veto and 0.635 points for the Russian veto. With covariates adjustment, the drop in *support*

¹⁶ I will create two models: one that includes the pre-treatment covariates and one that does not.

score is 0.617 points for the French veto and 0.661 points for the Russian veto. Both b_1 and b_2 are significant at the 0.1 percent level, which confirms that $Y(Y, Y) > Y(Y, N)$ and $Y(Y, Y) > Y(N, Y)$. However, when I conducted an one-way t-test to determine if $b_1 > b_2$ (with and without covariates adjustment), I fail to reject the null. Therefore, the Russian veto decreased support for the U.S. proposed war as much as the French veto, contrary to the Crawford-Sobel model. Finally, I conduct t-tests to determine whether $b_2 > b_1 + b_3$ and $0 > b_2 + b_3$. In both cases, I reject the null, which suggests that *support scores* in the two vetoes condition are the lowest of the four groups. The ordination in Hypothesis 1 is $Y(Y, Y) > Y(Y, N) > Y(N, Y) > Y(N, N)$; in contrast, my data suggest that $Y(Y, Y) > Y(Y, N) = Y(N, Y) > Y(N, N)$.

Table 16: Effect of treatments on Support Scores
Without pre-treatment covariates
Dependent variable: support score (1=strongly oppose, 5=strongly support)

	Coef.	Robust SE	t	P-value
France NO	-0.622	0.109	-5.720	0.000
Russia NO	-0.635	0.104	-6.090	0.000
France NO*Russia NO	0.357	0.154	2.310	0.021
Constant	3.750	0.070	53.640	0.000

N = 762
F(3, 758) = 27.52
Prob > F = 0.0000
R-squared = 0.0896

Table 17: Effect of treatments on support scores
 With pre-treatment covariates
 Dependent variable: support score (1=strongly oppose, 5=strongly support)

	Coef.	Robust SE	t	P-value
France NO	-0.613	0.109	-5.620	0.000
Russia NO	-0.657	0.102	-6.440	0.000
France NO*Russia NO	0.361	0.153	2.370	0.018
Male	0.183	0.076	2.420	0.016
Conservative	0.164	0.035	4.640	0.000
College	-0.099	0.076	-1.300	0.195
White	-0.114	0.103	-1.110	0.265
Democrats	0.221	0.111	1.980	0.048
Constant	3.121	0.221	14.100	0.000

Number = 761
 F(8, 752) = 14.45
 Prob > F = 0.0000
 R-squared = 0.1320

My data confirm existing survey experiments' findings that military campaigns approved by international organizations experienced higher support among the public (Grieco et al. 2011, Tingley & Tomz 2012). Nevertheless, the data show that subjects considered the Russian veto to be informative enough to change their attitudes. As an implicit manipulation check of my treatments, I asked my respondents to select from a list countries that they think are U.S. military allies. Sixteen percent of respondents selected Russia and 75 percent of respondents selected France. Likewise, I asked my respondents to select from a list countries that they think threaten U.S. national security. Thirty-five percent of respondents selected Russia and 1 percent of respondents selected France. Thus, in general, subjects were aware that France and the U.S. share greater foreign policy preferences than Russia and the U.S., which is one of my main assumptions.

Why did the Russia veto, which could be viewed as uninformative cheap talk, deter the Americans from supporting the hypothetical war? Tingley and Walter's (2011) lab experiment, targets were more likely to back down if they received a cheap talk threat

in an entry-deterrence game. The authors provide three reasons for this seemingly irrational behavior. First, cheap talk was especially influential in early rounds of play when challengers and defenders alike have very little information. Therefore, challengers act on threats, however costless, because these are the only hints about type. Second, the formal model assumes that defenders and challengers have common knowledge of the game, but in reality, the subjects might have not understood the logic of the game. Third, deviations from the model could be explained by subjects' heterogeneous psychology.

These explanations could also explain why subjects in my survey experiment expressed lower support for war because the vignette is analogous to a one-shot deterrence game. Although subjects had information about the "defender's" type, they still could not differentiate between informative cheap talk (French veto) and uninformative cheap talk (Russian veto). Nevertheless, France and Russia are not merely uni-dimensional types (i.e., ally versus rival). For instance, for some Americans, France is perceived to be weak while Russia is perceived to be strong. Advising George W. Bush in 2003, Condoleezza Rice said, "punish France, ignore Germany and forgive Russia."¹⁷ Rice's statement suggests that the U.S. is more powerful than France but not so powerful that it can allow prolonged hostility with Russia to continue. If subjects perceive Russia to be a powerful type and French to be a weak type, then they may take the Russian veto seriously as a deterrence.

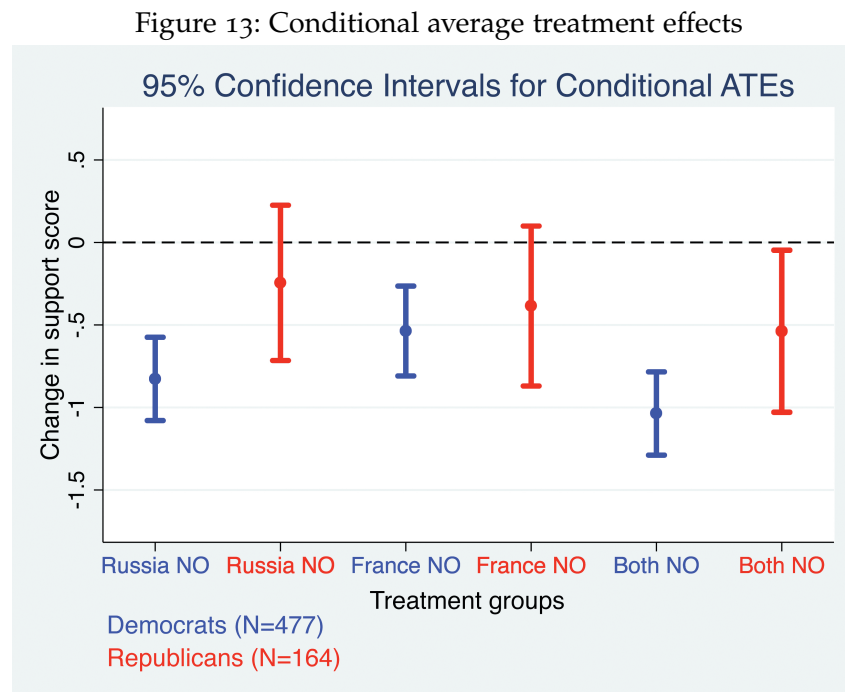
4.4.3 *Heterogeneous treatment effects*

In my second hypothesis, I state that foreign opposition to war will have a greater effect on Democrats' support for war than on Republicans' support for war. I define CATE as $E[Y_i(\text{Treatment}_n) - Y_i(\text{Control})]$, where n is the treatment number (1, 2, or 3). I estimate the CATEs using randomization inference and linear regression.

I conduct randomization tests using the R randomization inference package `ri` to estimate the CATE of each treatment. Samii and Aronow (2012) demonstrate that randomization-based constant effects estimator is equivalent to the OLS estimator using White's "robust"

¹⁷ Beale, Jonathan. "Can Rice's trip close rift with Europe." *BBC News*. February 3, 2005.

standard errors. Figure 13 shows the 95 percent confidence intervals for the CATEs. As Figure 13 shows, the drops in support for war are much bigger for Democrats and Republicans. In fact, for Republicans in Treatment 1 (the Russian veto) and Treatment 2 (the French veto), I failed to reject the sharp null of no treatment effect. For Republicans in Treatment 3, the treatment effect is barely significant. In contrast, all Democrats in all three treatment groups experienced significant treatment effects. Furthermore, the variance in Republicans' responses is much higher than the variance in Democrats' responses. While the Republicans' *support scores* somewhat observed the model in Hypothesis 1, Democrats' did not. In fact, for Democrats, the drop in *support score* is somewhat bigger in Treatment 2 (-0.245) than in Treatment 1 (-0.385).



Next, I estimate the CATEs using linear regressions. To test Hypothesis 2, I run the regression model from the previous subsection separately for Republicans and for Democrats. Then I compare the CATEs between the Republicans and Democrats. If my hypothesis is correct, then $b_{\text{Democrats}1} > b_{\text{Republicans}1}$, $b_{\text{Democrats}2} > b_{\text{Republicans}2}$, and

$(b_{\text{Democrats1}} + b_{\text{Democrats2}} + b_{\text{Democrats3}}) > (b_{\text{Republicans1}} + b_{\text{Republicans2}} + b_{\text{Republicans3}})$.
 The results from the regressions are reported in Table 18 and 19. Confirming the estimates from the randomization inference, the CATEs of the Republicans are much smaller than the CATEs of the Democrats. When I perform χ^2 test on the coefficients, I discover that there is no significant difference between Democrats and Republicans' responses to the French veto ($\chi^2=2.28$, p-value=0.1313), meaning $b_{\text{Democrats2}} = b_{\text{Republicans2}}$. On the other hand, there exists a significant difference, at the 5 percent level, between Democrats and Republicans' response to the Russian veto ($\chi^2=4.42$, p-value=0.0356). This finding suggests that $b_{\text{Democrats1}} > b_{\text{Republicans1}}$. Finally, there exists no significant difference between Democrats and Republicans' response to the two-vetoes treatment, meaning $(b_{\text{Democrats1}} + b_{\text{Democrats2}} + b_{\text{Democrats3}}) = (b_{\text{Republicans1}} + b_{\text{Republicans2}} + b_{\text{Republicans3}})$ ($\chi^2=2.29$, p-value=0.1304). The differences between Republicans' and Democrats' support scores in each treatment might be underestimated because I could not include robust standard errors in the χ^2 test for inequality between coefficients.

Table 18: Democrats: effect of treatments on support scores
 Without pre-treatment covariates
 Dependent variable: support score (1=strongly oppose, 5=strongly support)

	Coef.	Robust SE	t	P-value
France NO	-0.811	0.131	-6.190	0.000
Russia NO	-0.838	0.119	-7.060	0.000
France NO*Russia NO	0.601	0.184	3.270	0.001
Constant	3.829	0.077	49.590	0.000
Number of obs = 463				
F(3, 459) = 31.58				
Prob > F = 0.0000				
R-squared = 0.1490				

Table 19: Republicans: effect of treatments on support scores
 Without pre-treatment covariates
 Dependent variable: support score (1=strongly oppose, 5=strongly support)

	Coef.	Robust SE	t	P-value
France NO	-0.386	0.252	-1.530	0.127
Russia NO	-0.297	0.231	-1.290	0.200
France NO*Russia NO	0.012	0.349	0.030	0.974
Constant	3.872	0.161	24.080	0.000

Number of obs = 161
 F(3, 157) = 2.69
 Prob > F = 0.0480
 R-squared = 0.0453

My data show that Democrats and Republicans reacted more alike to the French veto than to the Russian veto. One interpretation is that the French veto sends a similar signal to both Republicans and Democrats while the Russian veto does not. For Democrats, the Russian veto depressed support for war even more than the French veto. In contrast, the Russian veto did little to sway Republicans' opinions. In fact, the Russian veto induced some Republicans to express higher support for war. It may be analogous to the "Freedom Fries" phenomenon in 2003 when the Republicans ignored and critiqued foreign opposition while rallying around the Iraq War. Interestingly, I do not observe such behavior among my Republican survey subjects: the mean *support score* in Treatment 2 (the French veto) is about the same for both parties.

Another chief difference between Republicans and Democrats is the variance in their *support scores*. Looking at Figure 13, the 95 percent confidence intervals for Democrats are about the same length — around 0.5 points. In contrast, the 95 percent confidence intervals for Republicans are much greater – around 1.0 points. Even if on average Republicans' *support scores* are negative in Treatment 1 and 2, the high variance in responses make it impossible to reject the sharp null of no effect. This data suggests that, in general, Democrats are more unified in responding to foreign opposition to U.S.-proposed war than Republi-

cans. This is somewhat surprising because in 2002–03, when foreign opposition to the Iraq War became progressively vocal, Republicans collectively supported the war (Hayes & Guardino 2010). Nevertheless, responses from Democratic subjects show that Democrats can be swayed by foreign opposition to war, despite domestic elites' support for war, a phenomenon observed in Feldman, Huddy, and Marcus (2007) and Jacobson (2007).

4.5 CONCLUSION

On a superficial level, the results of my survey experiment confirm two central arguments in American politics: 1) voters' attitudes are somewhat incoherent (Converse 2006) and 2) voters have different policy preferences based on their partisan identity (Green, Palmquist & Schickler 2002). First, subjects' *support scores* did not match the predictions from the Crawford-Sobel model of strategic communication. Although the Russian veto is less informative than the French veto, both treatments depressed support for war by roughly the same amount. Second, subjects' response to the treatments varied somewhat along partisan lines. In general, support for war was higher among Republicans than Democrats when treated with the foreign opposition message.

This paper takes a first cut at the question at hand. Having confirmed that foreign opposition to war indeed depresses domestic support for war, I plan to study the mechanisms behind the changes in subjects' attitudes when they are treated with the veto messages. One common way to study mechanisms is through multiple-choice mediation questions. Unfortunately, the selection of the mediation questions and their multiple-choice answers requires assumptions that might not be true.¹⁸ In my survey experiment, I adopted a different approach: I simply asked subjects why they selected the answer choice they selected. Using these open-ended responses, I hope to construct topic models to estimate differences between each treatment groups' justification for their answer choices (Roberts et al. 2013). This method is not by no means foolproof because subjects can provide false or incomplete information in these open-ended responses. Nevertheless, I hope the open-

¹⁸ For critiques of mediation analysis, see Glynn (2012); Green, Ha and Bullock (2010); Bullock and Ha (2010).

ended response did not compel subjects to consider mechanisms that were not a part of their original reasoning.

Future research could improve among the design of my survey experiment by including all possible combinations of UNSC permanent UNSC member votes. For instance, one could test a vignette in which Russia and China vetoed and France and Britain did not. Employing more combinations of UNSC votes would be a boon for construct validity. The responses in my survey experiment might have resulted from subjects' attitudes towards France and Russia, not from attitudes towards American allies and rivals in general. By expanding the list of country names, we can ensure the treatment effect we detect is not due to individual country's characteristics or reputation.¹⁹

Ultimately, this paper cannot explain the complex relationship between foreign elites and the American public in the lead-up to the Iraq War. Nevertheless, it confirms the argument scholars and journalists have made that certain members of the American public were influenced by official foreign voices to oppose the Iraq War. My findings suggest that in our globalized society, elite cues not only come from domestic politicians and media sources but also from international organizations and heads of other states.

4.6 ACKNOWLEDGMENTS

I thank Allan Dafoe, Nikolay Marinov, Yuke Li, Peter Aronow, David Broockman, Alan Gerber, Paul Sniderman, and Michael Tomz for their helpful comments. Also I am grateful for comments from my classmates in "Reputation, Perception, and War."

¹⁹ Nevertheless, we should ensure the combinations of votes are somewhat plausible. For instance, a veto by the UK might require a sketch of the imagination.

Part III

CONCLUSION

5 INTERNATIONAL RELATIONS AND SURVEY EXPERIMENTS

I began this senior thesis with two goals in mind. First, I sought to understand why Americans decide to support or oppose military interventions. Second, I attempted to untangle the problem of survey experiment vignettes that cause confounding. Through three sets of survey experiments I made both substantive and methodological contributions to the field of experimental IR.

5.1 SUBSTANTIVE CONTRIBUTIONS

My survey experiments contributed to existing IR literature in three ways. In Chapter 2, I replicate and expand upon previous experimental studies of the democratic peace. Since Kant's (Kant 1795) essay *To Perpetual Peace: A Philosophical Sketch*, scholars have debated the existence and mechanisms of the democratic peace. Proponents of the theory have observed that democracies have gone to war against other democracies less frequently than against autocracies (Dafoe 2011, Gartzke 1998, Kacowicz 1995, Lemke & Reed 1996, Maoz & Abdolali 1989, Maoz & Russett 1993, Oneal & Russett 1999b, Ray 1995, Rousseau et al. 1996, Russett 1993, Russett & Oneal 2001, Russett, Oneal & Davis 1998, Signorino & Ritter 1999, Melvin & Singer 1976, Thompson & Tucker 1997). Opponents, however, argue that factors other than representative institutions (e.g., economic development, trade, or culture) are actually the real forces behind the democratic peace (Gartzke 2007, Mousseau 2000, Mousseau, Hegre & Oneal 2003). The results of my survey experiment does not confirm nor deny the validity of the theory. While the naïve vignettes and the vignettes with controls failed to lower subjects' support for war against the democratic aggressor, the natural experiment embed vignettes produced a significant effect. Because these mini-experiments are under-powered, I cannot conclude the data proves or disproves the exist-

tence of the democratic peace in a public opinion context. Nevertheless, my results suggest that Americans are more willing to go to war against an autocracy than a democracy when they are clearly aware of the aggressor country's political institutions.

In Chapter 3, I test whether UNSC endorsements increase public support for different types of military interventions and economic sanctions. In line with recent scholarship (Grieco et al. 2011, Tingley & Tomz 2012), I find endorsements by international organizations serve as elite cues for domestic audiences. In my survey experiment, UN approval increased subjects' support for military intervention and economic sanctions in response to both humanitarian crisis and security threats. Conditional average treatment effect is especially large for subjects who strongly support the UN.

In Chapter 4, I examine whether foreign opposition to war lowers public support for war, a finding documented in observational studies (Hayes & Guardino 2011). In my survey experiment, I not only test the effect of foreign opposition on public opinion but also the *variations* in foreign opposition. By the logic of Crawford-Sobel's model of strategic communication, opposition by allies should be more informative and credible than opposition by rivals. Therefore, I predict, a UN Security Council (UNSC) veto by an American ally (France) will decrease support for war more than a veto by an American rival (Russia). In addition, I hypothesize Democrats will express greater objection to a war than Republicans when there exists foreign opposition to the war. Results from my survey experiment fail to confirm my first hypothesis; instead, the French veto and the Russian veto lowered public support for war by the same amount. On the other hand, I find that Democrats, more than Republicans, were more persuaded by foreign opposition to the war — just as my second hypothesis suggests.

5.2 METHODOLOGICAL CONTRIBUTIONS

My senior thesis also contributes to the development and design of survey experiments. In the introduction, I identify a problem associated with survey experiment vignettes. Poorly

designed vignettes cause confounding because they affect subjects' beliefs not only along the dimension of interest but also along other dimensions. For instance, in a naïve design of the democratic peace experiment, the label of "autocracy" could imply the aggressor country is economically underdeveloped or located in an undesirable region of the world. I explicate the problem of confounding treatment vignettes using the concept of construct validity, DAGs, and instrumental variable methods. To overcome the problem of confounding in naïve designs, I proposed two alternatives. The first involves adding additional information to the vignettes to control for confounders. The second involves inventing a "natural experiment" and inserting it into the vignettes.

In Chapter 2, I conduct three mini-experiments about democratic peace using the three types of vignettes. Using placebo tests, I compare the amount of confounding caused by each vignette type. While the naïve vignettes did not pass a few of the placebo test questions, they fared better than I predicted. This finding suggests that subjects are considering a subset of democratic countries when they read the naïve vignettes. If subjects have strong priors about how the world works, they would guess the democratic aggressor country is a democracy in name only. Therefore, experimenters should either use plausible scenario types in their vignettes or conservatively interpret findings from naïve designs.

In Chapter 3, I make another methodological contribution by testing whether IR experimenters should use specific country names or a hypothetical country in their vignettes. In my survey experiment, I used a list of randomly rotating country names in one condition and a hypothetical, unnamed country in the other. My results suggest that the two methods are equivalent. Rotating through a list of plausible country names might prove useful for experimenters who want to create realistic scenarios *and* preserve external validity.

5.3 AREAS OF FUTURE RESEARCH

The study of public opinion and international relations, along with improvements to survey experiment design, require additional research. Drawing from the lessons I learned while writing this senior thesis, I recommend two areas for future research.

First, the problem of confounding vignettes has not been entirely resolved. Naïve vignettes in other survey experiments might also cause confounding. One example is Hainmueller and Hiscox's (2010) well-cited survey experiment of attitudes towards high-skilled versus low-skilled immigration. Inconsistent with economic models, they find that low-skilled and highly skilled natives both prefer highly skilled immigrants over low-skilled immigrants. Somewhat consistent with the fiscal burden models, they find that poor natives are more opposed to low-skilled immigration in states with high fiscal exposure. To that end, they propose that opposition to immigration might be influenced by racial factors. Indeed, the question wording Hainmueller and Hiscox assign, "highly skilled" versus "low-skilled," are likely to be confounded with race and ethnicity. In the U.S., the majority of low-skilled immigrants are Hispanic and the majority of high-skilled immigrants are Chinese or Indian. An alternative explanation for this study could be that American "natives" prefer Chinese and Indian immigrants to Hispanic ones.

Other scholars have attempted to control for confounds, such as race and ethnicity, by using conjoint analysis to test which attributes of immigrants subjects favor or disfavor (Hainmueller & Hopkins 2012, Iyengar 2012, Iyengar et al. 2012). Subjects, playing the role of immigration officers, were told to choose between two immigrants varying along many attributes. While these new studies are more unbiased in their estimates, they can only detect opposition to immigration on an *individual* level. Iyengar (2012) has shown that survey respondents are more sympathetic about allowing individual immigrants into the U.S. than allowing large number of immigrants into the U.S. Therefore, Americans' attitudes towards individual immigrants is not a good proxy for their attitudes towards immigra-

tion policy. To this end, scholars should address the problem of confounding vignettes to accurately measure attitudes towards immigration policy.

Another area for future research concerns drawing inference from open-ended responses. Responses to multiple-choice mediation questions are easy to analyze but they force researchers to make assumptions about mechanisms. Open-ended responses are not a panacea; subjects might lie or not reveal their true feelings. Nevertheless, these short sentences provide researchers with a window into the minds of the subjects. Skimming through open-ended responses in my survey experiments, I began to realize that subjects' reasoning is very different from what I predicted. For instance, in the democratic peace survey, many subjects failed to mention anything about the regime type of the aggressor country. Others provided very complex arguments explaining why they selected the answer choice they did. Open-ended responses can also help experimenters know if their treatment vignettes are implausible and offend subjects' sense of reality. Drawing inference from text is not as simple as running a program in Python or R. Using a combination of human coding, automated content analysis, and artful interpretation, experimenters may gain a better understanding of their subjects' thought process.

Part IV

APPENDIX

A APPENDIX A: TEXTS OF THE SURVEYS

Appendix A includes the texts of the three survey experiments conducted as part of my senior thesis. They include:

1. The Democratic Peace Experiment
2. The UN Security Council Endorsement Experiment: Due to the extensive length of the survey, I did not include it in Appendix A. Please see my website pantheon.yale.edu/~bz44/ for a copy of the text
3. The Foreign Opposition to War Experiment

Beginning

INFORMED CONSENT (IRB Policy 200)

You will be asked to complete an online research survey that will take approximately 7 minutes. This survey is part of a research study conducted by researchers at Yale University. The goal of this survey is to ask you some questions about yourself and obtain your views.

Findings from this study may be reported in scholarly journals, at academic seminars, and at research association meetings. The data will be stored at a secured location and retained indefinitely. No identifying information about you will be made public and any views you express will be kept completely anonymous. Your participation is voluntary. You are free to stop the survey at any time without penalty.

There are no known risks associated with this study beyond those associated with everyday life. Although this study will not benefit you personally, we hope that our results will add to knowledge about different types of people and their opinions. Once you have completed the survey, you will be paid \$0.85.

To participate in the study, you must be at least 18 years old. If you have any questions, you can contact the researchers at yalesurveyresearch@gmail.com. If you have any questions about your rights as a research participant or concerns about the conduct of this study, you may contact the Yale University Human Subjects Committee, Box 208304, New Haven, CT 06520-8304, 203-785-4688, human.subjects@yale.edu.

- I agree to participate.
- I do not agree to participate.

What is your age?

What is your sex?

- Male
- Female
- Other

Are you a U.S. citizen?

- Yes
- No

Naive 1 dictator

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a dictator, who invades to get more power and resources.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Naive 2 democracy

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a democratically-elected government, which invades to get more power and resources.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Control1 developed dictator

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a dictator, who invades to get more power and resources. The attacking country is economically developed and is a major trading partner with the U.S.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Control 2 underdeveloped dictator

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a dictator, who invades to get more power and resources. The attacking country is economically underdeveloped and is a minor trading partner with the U.S.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Control3 developed democracy

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a democratically-elected government, which invades to get more power and resources. The attacking country is economically developed and is a major trading partner with the U.S.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
-

- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Control4 underdeveloped democracy

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

Select "I understand" to continue the survey

- I understand

A country sends its military to take over a neighboring country. The attacking country is led by a democratically-elected government, which invades to get more power and resources. The attacking country is economically underdeveloped and is a minor trading partner with the U.S.

Do you agree or disagree that the U.S. should use its military to stop the attacking country from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Natural experiment democracy

We are going to describe a hypothetical country, called Country A. We are also going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

You have 60 seconds to read the passage. Once the 60 seconds is up, the survey will automatically advance.

These page timer metrics will not be displayed to the recipient.

First Click: 0 seconds.

Last Click: 0 seconds.

Page Submit: 0 seconds.

Click Count: 0 clicks.

0100

You have 60 seconds to read this passage. Once the 60 seconds is up, the survey will automatically advance.

Five years ago, Country A was a fragile democracy. It had a democratically elected government, headed by a popular

president. At the time, a well-researched U.S. State Department report concluded that Country A was a democracy mainly because its president was very popular. The report also commented that without the president, there was a high probability that the country's military would overthrow the government to set up a dictatorship.

Two years ago at a public event, a disgruntled military officer shot at the president of Country A. The president was hit in the shoulder and survived the attack. Country A's democratically elected government survived the political turmoil, and is still a democracy today.

Recently, Country A sent its military to take over a neighboring country. The reason behind this invasion is to get more power and resources.

Do you agree or disagree that the U.S. should use its military to stop Country A from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Natural experiment dictator

We are going to describe a hypothetical country, called Country A. We are also going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds.*

Last Click: *0 seconds.*

Page Submit: *0 seconds.*

Click Count: *0 clicks.*

0100

You have 60 seconds to read the passage. Once the 60 seconds is up, the survey will automatically advance.

Five years ago, Country A was a fragile democracy. It had a democratically elected government, headed by a popular president. At the time, a well-researched U.S. State Department report concluded that Country A was a democracy mainly because its president was very popular. The report also commented that without the president, there was a high probability that the country's military would overthrow the government to set up a dictatorship.

Two years ago at a public event, a disgruntled military officer shot at the president of Country A. The president was hit in the head and did not survive the attack. In the political vacuum that followed the president's death, the country's military overthrew the democratically-elected government. Today, Country A is a military dictatorship.

Recently, Country A sent its military to take over a neighboring country. The reason behind this invasion is to get more power and resources.

Do you agree or disagree that the U.S. should use its military to stop Country A from invading its neighbor?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Open-ended response

Why did you select that answer choice in the previous question?

Placebo Test

How likely do you think it is that a scenario such as the one we just described could take place in an economically developed part of the world?

- Very likely
- Somewhat likely
- Undecided
- Somewhat unlikely
- Very unlikely

Of all the regions in the world, what regions do you think are most likely to experience a scenario such as the one we just described?

Pick the 3 most likely regions by dragging and dropping each item into the box. The regions will be ranked, with 1 being the most likely and 3 being the 3rd most likely.

Items

- Oceania
- Middle East and North Africa
- Eastern Europe
- Western Europe
- Central Asia
- East Asia
- North America
- South America
- Subsaharan Africa

The 3 most plausible regions

Central America and the
Caribbean
South Asia

Consider real world situations similar to the scenario we described. How likely do you think it is that the aggressor country in such a situation would violate the human rights of its citizens by imprisoning citizens because of their political beliefs?

- Very likely
- Somewhat likely
- Undecided
- Somewhat unlikely
- Very unlikely

Background info

What is the highest level of education you have completed?

- Less than high school
- High school
- Associate/Junior College
- Bachelor's
- Graduate's (Master's, MBA, PhD, MD)
- I don't know

Here is a 7-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?

	extremely liberal	liberal	slightly liberal	moderate, middle of the road	slightly conservative	conservative	extremely conservative
	1	2	3	4	5	6	7
Move the slider							

Are you of Spanish, Hispanic, or Latino descent?

- Yes, I am
- No, I am not
- I prefer not to say

Please check one or more categories below to indicate what race(s) you consider yourself to be.

- White Pacific Islander

Black or African American Some other race American Indian or Alaska Native I prefer not to say Asian

How often do you read news about foreign countries in newspapers or on the internet?

- Never
- Less than Once a Month
- Once a Month
- 2-3 Times a Month
- Once a Week
- 2-3 Times a Week
- Daily

How often do you watch news about foreign countries on television or on the internet?

- Never
- Less than Once a Month
- Once a Month
- 2-3 Times a Month
- Once a Week
- 2-3 Times a Week
- Daily

Besides English, how many other languages do you speak? If you speak one or more foreign languages, please write down which one(s).

- Zero
- One. I also speak
- Two. I also speak
- Three, I also speak
- Four or more, I also speak

Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

- Strong Democrat
- Weak Democrat
- Independent, leaning Democrat
- Independent
- Independent, leaning Republican
-

- Weak Republican
- Strong Republican
- Other

What is your Mechanical Turk ID number?

Debrief

Thank you for completing this survey!

You have been part of a randomized survey experiment to determine whether Americans are more likely to support military action against autocracies versus democracies. This theory in international relations is commonly known as the "democracy peace." To test this theory, subjects were randomly assigned to two treatment groups. Those in the treatment group read about a dictatorship invading its neighbor. Those in the control group read about a democracy invading its neighbor. Then we measured subjects' support for the U.S. military action against the attacking country for both the treatment and control groups.

Thank you for your participation. If you have any questions about the research, you can contact the researchers at yalesurveyresearch@gmail.com. If you have any questions about your rights as a research participant or concerns about the conduct of this study, you may contact the Yale University Human Subjects Committee at human.subjects@yale.edu.

The numerical code for this survey is 1812.

Default Question Block

INFORMED CONSENT (IRB Policy 200)

You will be asked to complete an online research survey that will take approximately 7 minutes. This survey is part of a research study conducted by researchers at Yale University. The goal of this survey is to ask you some questions about yourself and obtain your views.

Findings from this study may be reported in scholarly journals, at academic seminars, and at research association meetings. The data will be stored at a secured location and retained indefinitely. No identifying information about you will be made public and any views you express will be kept completely anonymous. Your participation is voluntary. You are free to stop the survey at any time without penalty.

There are no known risks associated with this study beyond those associated with everyday life. Although this study will not benefit you personally, we hope that our results will add to knowledge about different types of people and their opinions. Once you have completed the survey, you will be paid \$0.85.

To participate in the study, you must be at least 18 years old. If you have any questions, you can contact the researchers at yalesurveyresearch@gmail.com. If you have any questions about your rights as a research participant or concerns about the conduct of this study, you may contact the Yale University Human Subjects Committee, Box 208304, New Haven, CT 06520-8304, 203-785-4688, human.subjects@yale.edu.

- I agree to participate.
- I do not agree to participate.

What is your Mechanical Turk Worker ID?

What is your age?

What is your sex?

- Male
- Female
- Other

Are you a United States citizen?

- Yes
- No

Pre-treatment Questions

Here is a 7-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?

	extremely liberal	liberal	slightly liberal	moderate, middle of the road	slightly conservative	conservative	extremely conservative
	1	2	3	4	5	6	7
Move the slider							

What is the highest level of education you have completed?

- Less than high school
 High school
 Associate's/Junior College
 Bachelor's
 Graduate's (Master's, MBA, PhD, MD)
 I don't know

Language Party

Besides English, how many other languages do you speak?

- 0
 1
 2
 3
 4 or more

Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

- Strong Democrat
 Weak Democrat
 Independent, leaning Democrat
 Independent
 Independent, leaning Republican
 Weak Republican
 Strong Republican
 Other

US allies

Of the following countries, which do you think have a formal military alliance with the U.S.? (You can check multiple answers.)

- Venezuela
- China
- Russia
- Pakistan
- France

- Egypt
- Brazil
- Iran
- United Kingdom
- Japan

U.S. rivals

Of the following countries, which do you think threaten U.S. national security interests? (You can check multiple answers.)

- | | |
|---|------------------------------------|
| <input type="checkbox"/> United Kingdom | <input type="checkbox"/> Egypt |
| <input type="checkbox"/> Russia | <input type="checkbox"/> Iran |
| <input type="checkbox"/> Brazil | <input type="checkbox"/> France |
| <input type="checkbox"/> China | <input type="checkbox"/> Venezuela |
| <input type="checkbox"/> Pakistan | <input type="checkbox"/> Japan |

Political knowledge battery

Below are 3 factual questions. You have 30 seconds to answer them.

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*.

Last Click: *0 seconds*.

Page Submit: *0 seconds*.

Click Count: *0 clicks*.

30

Which political party has a majority in the U.S. House of Representatives?

- The Democratic Party
- The Republican Party
- I don't know

What is the name of the current U.S. Secretary of State?

- John Kerry
- Hillary Clinton
- Condoleezza Rice
- Chuck Hagel
- I don't know

Who is the current prime minister of Great Britain?

- Nick Clegg
- Rupert Murdoch
- David Cameron
- Gordon Brown
- I don't know

UNSC passage

Below is a short passage about the United Nations Security Council (UNSC). We will ask you 3 quiz questions about this passage after you finish reading it. If you answer all of them correctly, you will receive an additional 25 cents in pay. You will not be able to go back to this passage when you are answering the 3 quiz questions.

ABOUT THE UN SECURITY COUNCIL (UNSC)

The U.N. Security Council has 5 permanent members, who have been on the Council since 1946. The **5 permanent members** are **the United States, the United Kingdom, France, China, and Russia**.

The Council also has **10 temporary members**, who are elected every two years by the U.N. General Assembly. This year, the temporary members are Azerbaijan, Argentina, Australia, Guatemala, South Korea, Morocco, Pakistan, Luxembourg, Rwanda, and Togo.

The Council sometimes passes resolutions about the use of military force. When deciding whether to pass a resolution, each of the 15 members gets one vote. **A resolution passes if at least 9 members vote YES. However, if any permanent member votes NO, the resolution fails. Thus, each of the 5 permanent members has the power to block or "veto" a resolution.**

- I understand

Now we will ask you some questions about the United Nations Security Council (UNSC). If you answer all of them correctly, you will receive an additional 25 cents in pay. You have 40 seconds to answer the 3 questions.

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*.

Last Click: *0 seconds*.

Page Submit: *0 seconds*.

Click Count: *0 clicks*.

40

Which 5 countries are permanent members of the UNSC?

- United States, United Kingdom, China, Russia, and India
- United States, United Kingdom, China, Russia, and France
- United States, Russia, China, Germany, and France
- United States, China, Brazil, Russia, and Japan

What happens if 1 of the 5 permanent members votes NO on a resolution?

- The resolution fails.
- The resolution passes if enough other countries vote YES.

How many countries are temporary members of the UNSC?

- 5 countries
- 6 countries
- 8 countries
- 10 countries
- 12 countries

Here are the correct answers to the questions you just answered:

- The 5 permanent members of the UNSC are **United States, United Kingdom, China, Russia, and France.**
 - If 1 of the 5 permanent members votes NO on a resolution, **the resolution fails.**
 - **10 countries** are temporary members of the UNSC.
-

Instructions for experimental part

We are going to describe a situation the United States has faced in the past and could face in the future. Some parts of the description may strike you as important; other parts may seem unimportant. Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

All pass

A country recently sent its military to take over a neighboring country. The attacking country is led by a dictator, who invaded to get more power and resources. The attacking country has a powerful military, and the neighbor is too weak to defend itself. The U.S. president says the invasion is immoral and will hurt U.S. interests. He wants the U.S. military to push out the invaders.

The U.S. proposes a resolution to the United Nations Security Council (UNSC) to use military force against the aggressor country. **The UNSC unanimously passes the resolution. All five permanent members vote YES on it.**

Here are the votes of the five permanent members of the UNSC.

YES on the resolution: United States, United Kingdom, China, France, Russia

NO on the resolution: None

Do you agree or disagree that the U.S. should use military force to push out the invaders?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

I don't know

Why did you select that answer choice in the previous question?

France and Russia veto

A country recently sent its military to take over a neighboring country. The attacking country is led by a dictator, who invaded to get more power and resources. The attacking country has a powerful military, and the neighbor is too weak to defend itself. The U.S. president says the invasion is immoral and will hurt U.S. interests. He wants the U.S. military to push out the invaders.

The United States proposes a resolution to the United Nations Security Council (UNSC) to use military force against the aggressor country. **The UNSC fails to pass the resolution because two permanent members of the UNSC vote NO on it.**

Here are the votes of the five permanent members of the UNSC.

YES on the resolution: United States, United Kingdom, China

NO on the resolution: France, Russia

Do you agree or disagree that the U.S. should use military force to push out the invaders?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Why did you select that answer choice in the previous question?

France vetoes

A country recently sent its military to take over a neighboring country. The attacking country is led by a dictator, who invaded to get more power and resources. The attacking country has a powerful military, and the neighbor is too weak to defend itself. The U.S. president says the invasion is immoral and will hurt U.S. interests. He wants the U.S. military to push out the invaders.

The United States proposes a resolution to the United Nations Security Council (UNSC) to use military force against the aggressor country. **The UNSC fails to pass the resolution because one permanent member of the UNSC votes NO on it.**

Here are the votes of the five permanent members of the UNSC.

YES on the resolution: United States, United Kingdom, China, Russia

NO on the resolution: France

Do you agree or disagree that the U.S. should use military force to push out the invaders?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Why did you select that answer choice in the previous question?

Russia vetoes

A country recently sent its military to take over a neighboring country. The attacking country is led by a dictator, who invaded to get more power and resources. The attacking country has a powerful military, and the neighbor is too weak to defend itself. The U.S. president says the invasion is immoral and will hurt U.S. interests. He wants the U.S. military to push out the invaders.

The United States proposes a resolution to the United Nations Security Council (UNSC) to use military force against the aggressor country. **The UNSC fails to pass the resolution because one permanent member of the UNSC vetoes it.**

Here are the votes of the five permanent members of the UNSC.

YES on the resolution: United States, United Kingdom, China, France

NO on the resolution: Russia

Do you agree or disagree that the U.S. should use military force to push out the invaders?

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree
- I don't know

Why did you select that answer choice in the previous question?

Demographics questions

Now we will ask a few questions about yourself and your attitudes.

Are you of Spanish, Hispanic, or Latino descent?

- Yes, I am
- No, I am not
- I prefer not to say

Please check one or more categories below to indicate what race(s) you consider yourself to be.

- | | |
|---|---|
| <input type="checkbox"/> White | <input type="checkbox"/> Pacific Islander |
| <input type="checkbox"/> Black or African American | <input type="checkbox"/> Some other race <input style="width: 150px;" type="text"/> |
| <input type="checkbox"/> American Indian or Alaska Native | <input type="checkbox"/> I prefer not to say |
| <input type="checkbox"/> Asian | |

How often do you read news about foreign countries in newspapers or on the internet?

- Never
- Less than Once a Month
- Once a Month
- 2-3 Times a Month
- Once a Week
- 2-3 Times a Week
- Daily

How often do you watch news about foreign countries on television or on the internet?

- Never
- Less than Once a Month
- Once a Month
- 2-3 Times a Month
- Once a Week

2-3 Times a Week Daily

Debrief

You have been part of a randomized survey experiment to test whether UNSC vetoes affect American public support for war. In the survey experiment, we present subjects with a hypothetical international crises that resulted in a U.S. proposal for military intervention. Subjects are then randomly divided into four experimental groups. The first group (control group) is told that the UNSC unanimously passed a resolution approving military action. The second group is told the UNSC resolution failed to pass because Russia vetoed it. The third group is told the UNSC resolution failed to pass because France vetoed it. The fourth group is told the UNSC resolution failed to pass because both Russian and France vetoed it. Following the experimental vignettes, we measure subjects' support for the proposed war.

Thank you for your participation. If you have any questions about the research, you can contact the researchers at yalesurveyresearch@gmail.com. If you have any questions about your rights as a research participant or concerns about the conduct of this study, you may contact the Yale University Human Subjects Committee at human.subjects@yale.edu.

The numeric code for this survey is 1914.

B APPENDIX B: ADDITIONAL GRAPHS

Figure 14: Naïve vignette type: distribution of support scores

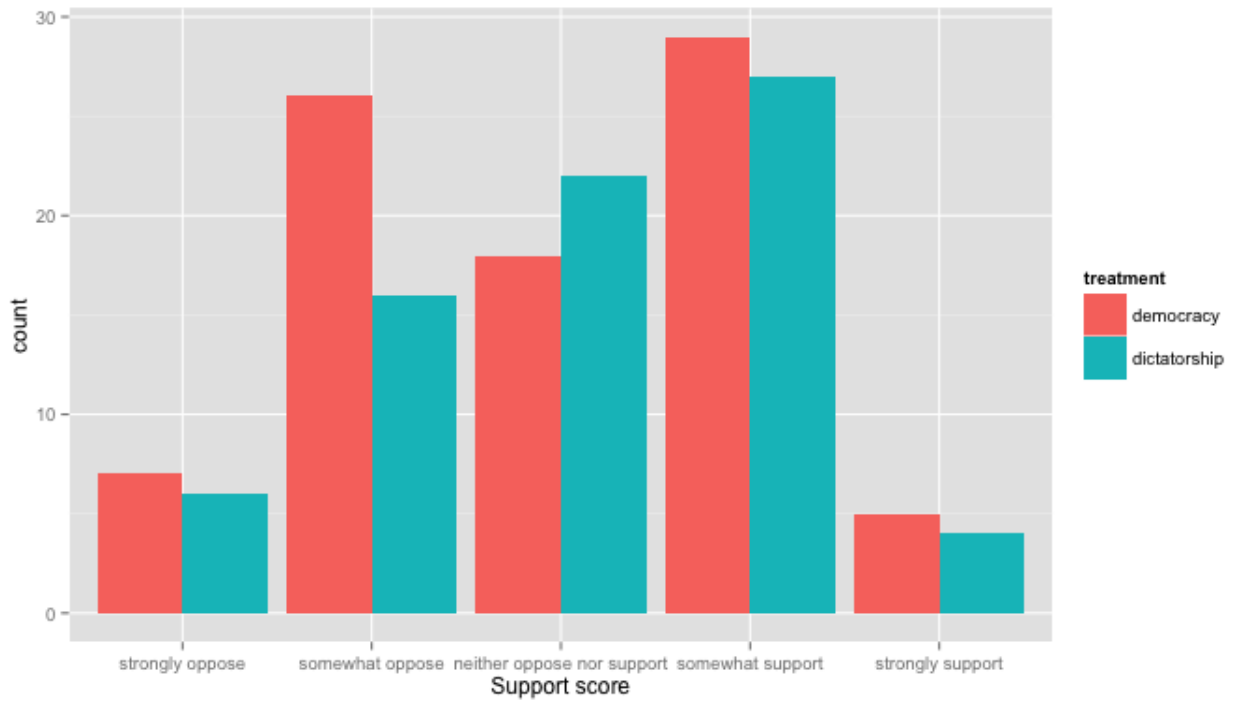


Figure 15: Controls vignette type: distribution of support scores

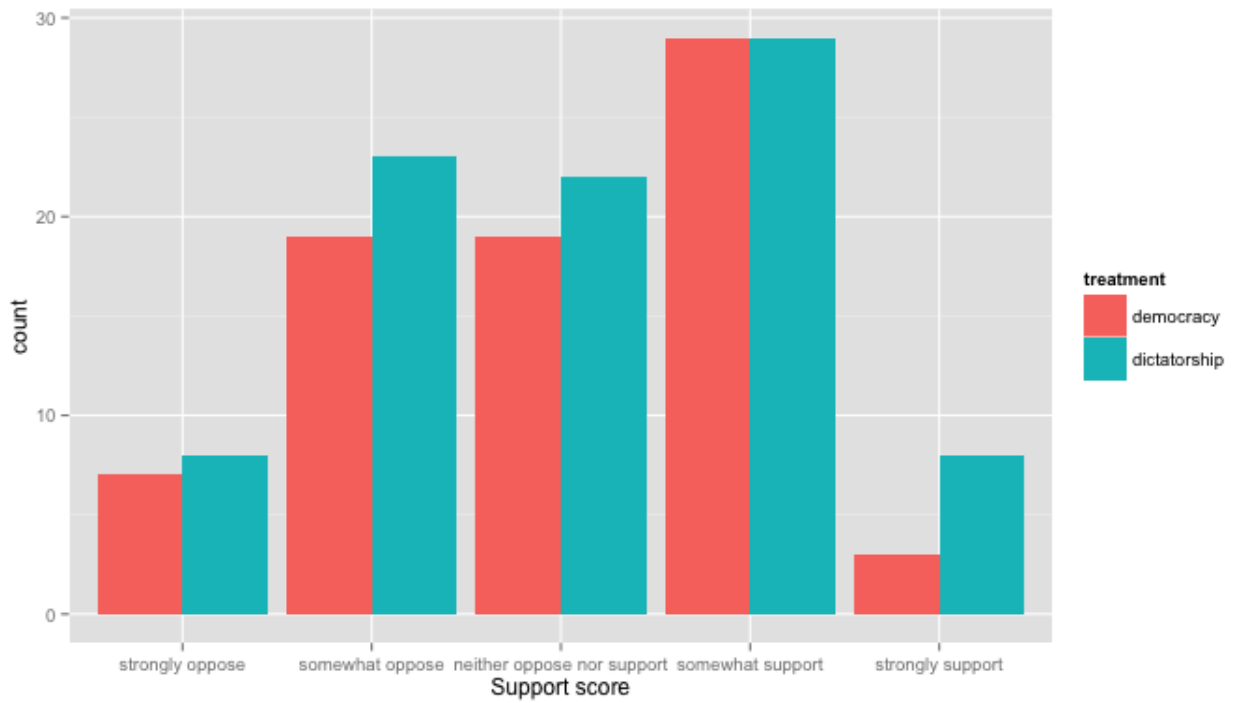
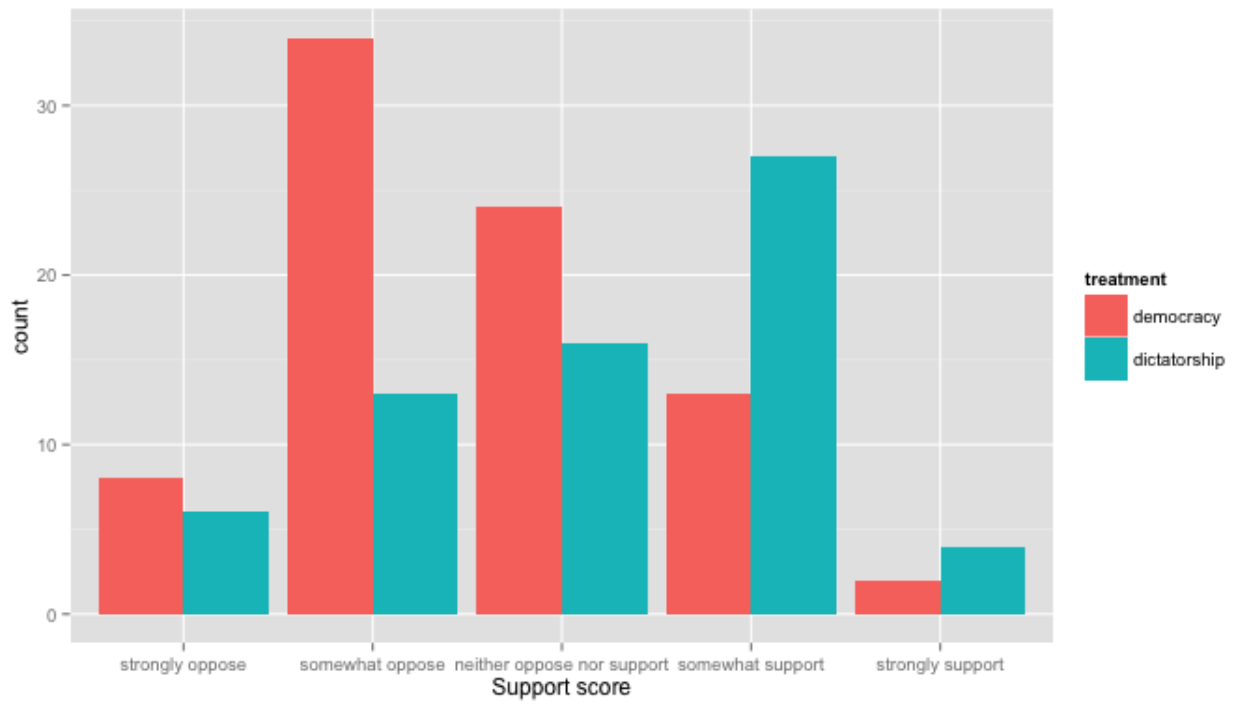


Figure 16: Natural experiment vignette type: distribution of support scores



BIBLIOGRAPHY

- Abbott, Kenneth. W. & Duncan Snidal. 1998. "Why states act through formal international organizations." *Journal of Conflict Resolution* 42(1):3–32.
- Aldrich, John A., John L. Sullivan & Eugene Borgida. 1989. "Foreign Affairs and Issue Voting: Do Presidential Candidates "Waltz Before A Blind Audience?"" *American Political Science Review* 83(1):123–141.
- Barabas, Jason & Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104(2):226–242.
- Beer, Francis A, Grant P Sinclair, Alice F Healy & Lyle E Bourne Jr. 1995. "Peace agreement, intractable conflict, escalation trajectory: A psychological laboratory experiment." *International Studies Quarterly* pp. 297–312.
- Berinsky, Adam., Greg Huber & Gabriel Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon. com's Mechanical Turk." *Political Analysis* 20(3):351–368.
- Berinsky, Adam J. 2007. "Assuming the costs of war: Events, elites, and American public support for military conflict." *Journal of Politics* 69(4):975–997.
- Berinsky, Adam J. 2009. *In Time of War: Understanding American Public Opinion from World War II to Iraq*. Chicago: University of Chicago Press.
- Brady, Henry E. 2000. "Contributions of Survey Research to Political Science." *PS: Political Science and Politics* 33(1):47–57.
- Bullock, John G & Shang E Ha. 2010. "Mediation analysis is harder than it looks." *Cambridge Handbook of Experimental Political Science* .
- Campbell, Angus, Philip Converse, Warren Miller & Donald Stokes. 1960. *The American Voter*. New York: John Wiley and Son.
- Campbell, D. T. & J. C. Stanley. 1963. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Caughey, Devin, Allan Dafoe & Jason Seawright. 2013. "Testing Elaborate Theories in Political Science: Nonparametric Combination of Dependent Tests." Manuscript.
- Chapman, Terrence L. & Dan Reiter. 2004. "The United Nations Security Council and the rally 'round the flag effect." *Journal of Conflict Resolution* 48(6):886–909.
- Converse, Philip E. 2006. "The nature of belief systems in mass publics (1964)." *Critical Review* 18(1-3):1–74.

- Crawford, Vincent P. 1998. "A Survey of Experiments on Communication via Cheap Talk." *Journal of Economic Theory* 78(2):286–298.
- Crawford, Vincent P. & Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):1431–1451.
- Dafoe, Allan. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55(2):247–262.
- Deutsch, Morton, Yakov Epstein, Donnan Canavan & Peter Gumpert. 1967. "Strategies of Inducing Cooperation: An Experimental Study." *Journal of Conflict Resolution* 11(3):345–360.
- Druckman, James N. & Kjersten R. Nelson. 2003. "Framing and Deliberation: How Citizens' Conversations Limit Elite Influence." *American Journal of Political Science* 47(4):729–745.
- Drury, A. Cooper. 2001. "Sanctions as Coercive Diplomacy: The US President's Decision to Initiate Economic Sanctions." *Political Research Quarterly* 54(3):485–508.
- Drury, A. Cooper. 2005. *Economic sanctions and presidential decisions: Models of political rationality*. New York: Palgrave Macmillan.
- Fearon, James D. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88(03):577–592.
- Feldman, Stanley, Leonie Huddy & George Marcus. 2007. "Going to War: When Citizens Matter." Paper prepared for presentation at the fall session of the New York Area Political Psychology Meeting, Columbia University.
- Gaines, Brian J., James H. Kuklinski & Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15(1):1–20.
- Gartzke, Erik. 1998. "Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42(1):1–27.
- Gartzke, Erik. 2007. "The Capitalist Peace." *American Journal of Political Science* 51(1):166–191.
- Gelpi, C., P. Feaver & J.A. Reifler. 2009. *Paying the human costs of war: American public opinion and casualties in military conflicts*. Princeton, New Jersey: Princeton University Press.
- Gerring, John. 2001. *Party ideologies in America, 1828-1996*. Cambridge, UK: Cambridge University Press.
- Gibson, James L., Gregory A. Caldeira & Lester Kenyatta Spence. 2005. "Why Do People Accept Public Policies They Oppose? Testing Legitimacy Theory with a Survey-Based Experiment." *Political Research Quarterly* 58(2):187–201.
- Gilens, Martin. 2001. "Political Ignorance and Collective Policy Preferences." *American Journal of Political Science* 95(2):379–396.

- Gilens, Martin. 2002. *Navigating Public Opinion: Polls, Policy, and the Future of American Democracy*. Oxford: Oxford University Press chapter An Anatomy of Survey Based Experiments.
- Glynn, Adam N. 2012. "The Product and Difference Fallacies for Indirect Effects." *American Journal of Political Science* 56(1):257–269.
- Green, Donald P., Bradley Palmquist & Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven, CT: Yale University Press.
- Green, Donald P, Shang E Ha & John G Bullock. 2010. "Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose." *The ANNALS of the American Academy of Political and Social Science* 628(1):200–208.
- Grieco, Joseph. M., Christopher. Gelpi, Jason Reifler & Peter. D. Feaver. 2011. "Let's Get a Second Opinion: International Institutions and American Public Support for War." *International Studies Quarterly* 55(2):563–583.
- Hainmueller, Jens & Daniel J. Hopkins. 2012. "The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes Toward Immigrants." Manuscript.
- Hainmueller, Jens & Michael J. Hiscox. 2010. "Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review* 104(1):61–84.
- Hayes, Danny & Matt Guardino. 2010. "Whose views made the news? Media coverage and the march to war in Iraq." *Political Communication* 27(1):59–87.
- Hayes, Danny & Matt Guardino. 2011. "The Influence of Foreign Voices on US Public Opinion." *American Journal of Political Science* 55(4):831–851.
- Herrmann, Richard K., Philip E. Tetlock & Penny S. Visser. 1999. "Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93(3):553–573.
- Holsti, Ole Rudolf. 2004. *Public opinion and American foreign policy*. University of Michigan Press.
- Hurwitz, Jon & Mark Peffley. 1997. "Public Perceptions of Race and Crime: The Role of Racial Stereotypes." *American Journal of Political Science* 41(2):375–401.
- Iyengar, Shanto. 2012. "Do Attitudes About Immigration Predict Willingness to Admit Individual Immigrants? A Cross-National Test of the Person-Positivity Bias." Manuscript.
- Iyengar, Shanto, Nicholas A. Valentino, Ray Duch, Kru Hahn & Tetsuro Kobayashi. 2012. "Economic and Cultural Threat as Antecedents of Opposition to Immigration: Cross-National Evidence from the U.S., U.K., Japan, and South Korea." Manuscript.
- Jacobson, Gary C. 2007. *A Divider, Not a Uniter: George W. Bush and the American People*. New York: Pearson Longman.

- Jentleson, Bruce W. & Rebecca L. Britton. 1998. "Still Pretty Prudent Post-Cold War American Public Opinion on the Use of Military Force." *Journal of Conflict Resolution* 42(4):395-417.
- Jervis, Robert. 1989. *The logic of images in international relations*. Columbia University Press.
- Johns, Robert & Graeme A. M. Davies. 2012. "Democratic Peace or Clash of Civilizations? Target States and Support for War in Britain and the United States." *Journal of Politics* 74(4):1038-1052.
- Jones, Benjamin F. & Benjamin A. Olken. 2009. "Hit or Miss? The Effect of Assassinations on Insitutions and War." *American Economic Journal: Macroeconomics* 1(2):55-87.
- Kacowicz, Arie M. 1995. "Explaining Zones of Peace: Democracies as Satisfied Powers." *Journal of Peace Research* 32(3):265-276.
- Kant, Immanuel. 1795. *To Perpetual Peace: A Philosophical Sketch*. 2003 ed. New York: Hackett.
- Karol, David & Edward Miguel. 2007. "The Electoral Cost of War: Iraq Casualties and the 2004 U.S. Presidential Election." *Journal of Politics* 69(03):633-648.
- Kohut, Andrew & Bruce Stokes. 2007. *America against the world: How we are different and why we are disliked*. New York: St. Martin's Griffin.
- Lemke, Douglas & William Reed. 1996. "Regime Types and Status Quo Evaluations: Power Transition Theory and the Democratic Peace." *International Interactions* 22(2):143-164.
- Lenz, Gabriel S. 2012. *Follow the Leader?: How Voters Respond to Politicians' Policies and Performance*. Chicago: University of Chicago Press.
- Lupia, Arthur & Mathew D. McCubbins. 1998. *The democratic dilemma: Can citizens learn what they need to know?* Cambridge, UK: Cambridge University Press.
- Luskin, Robert C., James F. Fishkin & Roger Jowell. 2002. "Considered Opinions: Deliberative Polling in Britain." *British Journal of Political Science* 32:455-487.
- Majeski, Stephen J & Shane Fricks. 1995. "Conflict and cooperation in international relations." *Journal of Conflict Resolution* 39(4):622-645.
- Maoz, Zeev & Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87(3):624-638.
- Maoz, Zeev & Nasrin Abdolali. 1989. "Regime Types and International Conflict, 1816-1976." *Journal of Conflict Resolution* 33(1):3-35.
- Melvin, Small & J. David Singer. 1976. "The War-Proneness of Democratic Regimes." *Jerusalem Journal of International Relations* 1:50-68.
- Mintz, Alex & Nehemia Geva. 1993. "Why Don't Democracies Fight Each Other? An Experimental Study." *The Journal of Conflict Resolution* 37(3):484-503.

- Mintz, Alex, Yi Yang & Rose McDermott. 2011. "Experimental Approaches to International Relations." *International Studies Quarterly* 55:493–501.
- Mousseau, Michael. 2000. "Market Prosperity, Democratic Consolidation, and Democratic Peace." *Journal of Conflict Resolution* 44(4):472–507.
- Mousseau, Michael, Havard Hegre & John R. Oneal. 2003. "How the Wealth of Nations Conditions the Liberal Peace." *European Journal of International Relations* 9(2):277–314.
- Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton, New Jersey: Princeton University Press.
- Nincic, Miroslav & Barbara Hinckley. 1991. "Foreign Policy and the Evaluation of Presidential Candidates." *Journal of Conflict Resolution* 35(2):333–355.
- Oneal, John R. & Bruce Russett. 1999b. "Is the Liberal Peace Just an Artifact of Cold War Interests? Assessing Recent Critiques." *International Interactions* 25(3):213–241.
- Page, Benjamin I & Marshall M Bouton. 2006. *The foreign policy disconnect: What Americans want from our leaders but don't get*. University of Chicago Press.
- Pilisuk, Marc. 1984. "Experimenting with the arms race." *Journal of Conflict Resolution* 28(2):296–315.
- Putnam, Robert D. 1988. "Diplomacy and domestic politics: the logic of two-level games." *International Organization* 42(03):427–460.
- Ray, James Lee. 1995. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition*. Columbia, South Carolina: University of South Carolina Press.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson & David Rand. 2013. "Topic models for open ended survey responses with applications to experiments." Working paper.
- Roth, Alvan E. 1995. *The Handbook of Experimental Economics*. New Jersey, NJ: Princeton University Press chapter Bargaining Experiments.
- Rousseau, David L. 2005. *Democracy and War: Institutions, Norms, and the Evolution of International Conflict*. Stanford, California: Stanford University Press.
- Rousseau, David L., Christopher Gelpi, Dan Reiter & Paul K. Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918–1988." *American Political Science Review* 90(3):512–533.
- Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton, New Jersey: Princeton University Press.
- Russett, Bruce & John Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York, New York: W. W. Norton.

- Russett, Bruce, John R. Oneal & David R. Davis. 1998. "The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes." *International Organization* 52(3):441–467.
- Samii, Cyrus & Peter M. Aronow. 2012. "On equivalencies between design-based and regression-based variance estimators for randomized experiments." *Statistics and Probability Letters* 82(2):365 – 370.
- Sanbonmatsu, Kira. 2002. "Gender stereotypes and vote choice." *American Journal of Political Science* 46(1):20–34.
- Schildkraut, Deborah J. 2009. "The Dynamics of Public Opinion on Ethnic Profiling After 9/11: Results From a Survey Experiment." *American Behavioral Scientist* 53(1):61–79.
- Shadish, William R. & Thomas D. Cook. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Signorino, Curtis S. & Jeffrey M. Ritter. 1999. "Tau-b or Not Tau-b: Measuring Similarity of Foreign Policy Positions." *International Studies Quarterly* 43(1):115–144.
- Sniderman, Paul M & Edward G Carmines. 1997. *Reaching beyond race*. Harvard University Press.
- Sniderman, Paul M., Louk Hagendoorn & Markus Prior. 2004. "Predisposing Factors and Situational Triggers: Exclusionary Reactions to Immigrant Minorities." *American Political Science Review* 98(1):35–49.
- Spence, Michael. 1973. "Job market signaling." *The quarterly journal of Economics* 87(3):355–374.
- Thompson, A. 2006. "Coercion through IOs: The Security Council and the logic of information transmission." *International Organization* 60(1):1–34.
- Thompson, William R. & Richard Tucker. 1997. "A Tale of Two Democratic Peace Critiques." *Journal of Conflict Resolution* 41(3):428–454.
- Tingley, Dustin H. & Barbara F. Walter. 2011. "Can Cheap Talk Deter? An Experimental Analysis." *Journal of Conflict Resolution* 55(6):996–997.
- Tingley, Dustin & Michael Tomz. 2012. "How does the UN Security Council Influence Public Opinion?" Submitted.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4):821–840.
- Tomz, Michael & Jessica L. Weeks. 2012a. "An Experimental Investigation of the Democratic Peace." Unpublished.
- Tomz, Michael & Jessica L. Weeks. 2012b. "Human Rights, Democracy, and International Conflict." Unpublished.

- Trager, Robert F. & Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal of Political Science* 55(3):526–545.
- Voeten, Eric. 2005. "The Political Origins of the UN Security Council's Ability to Legitimize the Use of Force." *International Organization* 59(3):527–557.
- Wegner, Daniel M. 1994. "Ironic processes of mental control." *Psychological Review* 101(1):34.
- Wooldridge, Jeffrey M. 2001. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT press.
- Zaller, John R. 1992. *The nature and origins of mass opinion*. Cambridge university press.

COLOPHON

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*".

DECLARATION

I undertake that all the material presented for evaluation is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for evaluation.

*New Haven, Conn.,
April 19, 2013*

Baobao Zhang,
April 19, 2013