Pre-Analysis Plan for Political Persuasion on the Treaty on the Prohibition of Nuclear Weapons: Evidence from Parallel Survey Experiments in the United States and Japan

Jonathon Baron,*† Rebecca Gibbons,‡ Stephen Herzog†‡

August 4, 2019

Abstract

Abstract here.

1 Overview

How do governments approach and resist unfavorable emerging norms in international relations? What types of messages are most effective in confronting these norms? We propose to address these questions with parallel exploratory survey experiments in the United States and Japan on the influence of persuasive messaging regarding the Treaty on the Prohibition of Nuclear Weapons (TPNW). Both US and Japanese governments oppose the TPNW, despite historically strong public support for nuclear disarmament in both countries. The context of the TPNW therefore provides a policy-relevant context in which to examine how governments use persuasive messaging to resist emerging norms. Our primary theoretical focus regards how particular persuasive arguments regarding nuclear disarmament affect public attitudes; but we also speak to the broader literature on political persuasion, as well as recent work on political cues in international relations (IR).

In particular, we ask: 1.) whether attitudes toward nuclear disarmament are more effectively moved by group cues, or by substantive arguments; and, 2.) among the various substantive arguments in favor of the TPNW, and disarmament more generally, which messages yield

*The authors view their contributions as equal, and their names are listed in alphabetical order.

†PhD Candidate, Yale University Department of Political Science, 115 Prospect St., New Haven, CT 06511

‡Postdoctoral Research Fellow, Harvard University Project on Managing the Atom, Belfer Center for Science and International Affairs, 79 John F. Kennedy St., Cambridge, MA 02138

the strongest persuasive effects? To explore these questions, we assess the persuasive effects of substantive arguments *against* the TPNW. We focus on anti-TPNW messaging for two primary reasons. First, empirically, the political market is already saturated with *pro*-TPNW messaging from anti-nuclear groups. These messages are consistent with mainstream public opinion, which favors disarmament in both countries. Focusing on the persuasive effects of counter-attitudinal messages therefore provides us with greater opportunity to assess potential routes of attitude *change*, and to more effectively advance theoretical discussions in the fields of political behavior and IR. Second, the US and Japanese governments are both opposed to the TPNW in contrast to broad-spectrum public approval of disarmament policies in both countries. This divergence creates incentives for each state to employ persuasive strategies to bring mass attitudes in line with government policy. Focusing on the effects of anti-TPNW messaging allows us to examine how each government addresses emerging norms, since we can identify the potential effects of government communications strategies in each country and pinpoint how state messaging may seek to target citizens. As our survey will be conducted as debates regarding the TPNW continue, our analyses will allow us to make generalizable claims about persuasive effects of interest to real-world political phenomena.

The Pre-Analysis Plan is organized as follows. Section 2: Surveying Procedure provides an overview of our experimental design and presents treatments, outcome measures, and covariate measures. Section 3: Proposed Analyses describes the statistical models and inferential strategies we intend to employ in estimating effects and conducting statistical hypothesis testing; planned subgroup analyses; and practices regarding missing values, "Don't know" responses, and other exclusion criteria. Having provided our statistical models in Section 3, Section 4: Hypotheses provides our theoretical hypotheses. Finally, Section 5: Instruments provides both our US and Japanese survey instruments.

2 Surveying Procedure

The survey will be conducted with the Dynata surveying firm (formerly SSI), with separate samples in the United States and Japan. The US sample includes 1,200 subjects and is nationally representative, balanced on age, gender, ethnicity, and US census region. The Japanese sample is nationally representative, balanced on age, gender, and region. The complete survey instruments are included below in Section 5: Instruments.

2.1 Institutional Review Board (IRB) Exemption

All studies are exempt from IRB review, persuant to Yale University IRB Protocol #2000026191.

2.2 Experimental Design Overview

Each experiment is a five-arm trial, assessing the relative influence of group and substantive cues on public attitudes toward the TPNW. Our US survey will feature $N = 1,200$ respondents, and our Japanese survey will feature $N = 1,000$ respondents. The instruments were designed to be as similar as possible: The group-cue treatment will be identical in each, save minor changes in language resulting from translation; because our substantive-cue treatments employ exact text from representatives of each government, we have selected statements that are as similar in language and meaning as possible; nonetheless, some differences do exist between the texts.

After consenting to the survey and answering several pre-treatment covariate questions, each subject will be randomly assigned to one of five treatment arms. The five arms are as follows.

1. **Baseline (Control)**: basic information about the TPNW (provided to all respondents);
2. **Group Cue**: augmented replication of Kertzer and Zeitzoff (2017)'s treatment indicating that people like the respondent view the TPNW, without substantive rationales;[1]
3. **Security Cue**: substantive argument describing the TPNW as conflicting with the country's national security interests.
4. **Norms Cue**: substantive argument stating that the TPNW would erode the norms of nuclear disarmament established by the Non-Proliferation Treaty.
5. **Institutions Cue**: substantive argument explaining that the TPNW insufficiently addresses verification and inspections, and that it would invite cheating.

2.3 Outcome Measures

To assess treatment effects, we will include outcome measures of support for the TPNW and attitudes toward nuclear disarmament:

TPNW Approval: single-choice, binary response.

Disarmament Attitudes: seven-item matrix grid, with randomized row order; four-point, single-choice Likert responses (Strongly agree, Agree, Disagree, Strongly Disagree)

The precise treatment text is provided in Section 5: Instruments below.

[1] In fact, this treatment uses mild deception. As a result, our survey concludes with a debrief noting that deception may have been used, and explaining our purposes in so doing.

2.4 Planned Covariate Measures

Our instrument includes covariate questions regarding gender (male, female, other), age, three-point partisan identification (PID3; -1 for Democrat, 0 for Independent, and 1 for Republican), five-point ideology (-2 for Very liberal, to 2 for Very conservative), education, geographic region, and self-reported 2018 family income.

Analyses will be conducted both with and without covariates, as described in Section 3: Proposed Analyses.

3 Proposed Analyses

Our analysis will focus on comparisons between the persuasive effects of partisan cues, substantive cues, and their interaction; to do so, we intend to use ordinary least squares (OLS) regressions, with HC2 robust standard errors (computed using the `sandwich` package in R). We will fit models both with and without covariates. We also intend to perform exploratory subgroup analyses in order to consider the potential influences of ideology and other demographic traits, such as gender. These results can shed light on questions of attitude formation and persuasion in political science, as well as in the context of nuclear policy and nuclear disarmament, more specifically.

3.1 Planned Effect Measurement

We intend to estimate effects relative to the baseline of pure control using OLS regression with HC2 robust standard errors. We plan to estimate effects per Model 1 and Model 2 below. Model 1 is defined as

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 \mathit{Group\ Cue}_i + \beta_2 \mathit{Security\ Cue}_i + \\ &\beta_3 \mathit{NPT\ Cue}_i + \beta_4 \mathit{Verifiability\ Cue}_i + \epsilon_i, \text{ and} \end{aligned} \tag{1}$$

Model 2 is defined similarly as

$$\begin{aligned} Y_i = &\gamma_0 + \gamma_1 \mathit{Group\ Cue}_i + \gamma_2 \mathit{Security\ Cue}_i + \\ &\gamma_3 \mathit{NPT\ Cue}_i + \gamma_4 \mathit{Verifiability\ Cue}_i + \theta \mathbf{X}_i + \eta_i. \end{aligned} \tag{2}$$

The sole difference between Model 1 and Model 2 is the inclusion of covariates in Model 2, where $\mathbf{X}_i$ represents the vector of covariates recorded for each respondent i (with θ representing the associated vector of estimated coefficients). ϵ_i and η_i represent the respective error terms.

We are interested in comparing the effects of each cue on outcome measures Y_i. Each model assesses the four cues in comparison to the baseline (control) condition. Therefore, β_1 and γ_1 represent estimates of the group-cue effects, both with and without covariates; β_2 and γ_2, the effects of the national-security cue in each model; β_3 and γ_3, the effects of the the norms cue; and, β_4 and γ_4, the effects of the institutional cue.

Note that we may also conduct analyses using additional or alternative statistical models. We may also conduct analyses with additional conditioning sets, e.g., including subsets of covariates included in $\mathbf{X}_i$.

3.2 Planned Inferential Methods (Statistical Hypothesis Testing)

Since we expect directional effects from each treatment, hypothesis testing for all main effects will be conducted against the null hypothesis using one-tailed p-values. For comparisons *between* effects, we will employ two-tailed p-values. We set $\alpha = .05$ for all inferential analyses.

3.3 Planned Subgroup Analyses

Subgroup analyses will center on demographic classifications of interest based on the existing literature. These classifications include gender, three-point partisan identification (PID3), five-point ideology, education, age, geographic region, and self-reported family income. As we do not provide specific hypotheses regarding subgroup analyses, these are technically exploratory.

3.4 Missingness

Where necessary, missing covariate values will be replaced with mean-imputed values, with means computed across all experimental conditions.

3.5 "Don't know" (DK) and Other Responses

Any skipped responses will be converted to `NA` values, and replaced with mean imputation. Missing values will also be converted to `NA` and replaced with mean imputation: 1.) when a respondent answers the initial ideology question with "Haven't thought much about this," and skips the follow-up ("If you had to choose, would you consider yourself a liberal or a conservative?"); 2.) when a respondent reports a family income of $0.

Two demographic variables—gender and education—allow "Other" responses. Where a respondent selects "Other" for either of these questions, provided text responses will be recorded in a separate variable; numerically, "Other" responses will be recoded as `NA` values,

and replaced with mean imputation. All other DK values will be coded to neutral on the relevant response scales.

3.6 Subject Exclusion Criteria

Each survey can only be completed by adult (over the age of 18) citizens of each respective country, who provide consent. There are no additional anticipated criteria for subject exclusion.

4 Hypotheses

We expect each treatment to have negative effects, yielding four straightforward and simple hypotheses:

Hypothesis 1: The group cue will have a negative effect on support for TPNW in both models ($\beta_1 < 0$ and $\gamma_1 < 0$).

Hypothesis 2: The security cue will have a negative effect on support for TPNW in both models ($\beta_2 < 0$ and $\gamma_2 < 0$).

Hypothesis 3: The norms cue will have a negative effect on support for TPNW in both models ($\beta_3 < 0$ and $\gamma_3 < 0$).

Hypothesis 4: The institutions cue will have a negative effect on support for TPNW in both models ($\beta_4 < 0$ and $\gamma_4 < 0$).

As noted in Section 3.2, all hypothesis testing will be conducted using one-tailed (lower) p-values, since we expect all effects to be negative, compared to baseline.

5 Instruments

5.1 US Instrument

<US survey text>

5.2 Japanese Instrument

〈日本語の調査〉

References

Kertzer, J. D. and T. Zeitzoff (2017). A bottom-up theory of public opinion about foreign policy. *American Journal of Political Science 61*(3), 543–558.