

Pre-Analysis Plan
Entrenched versus Shallow Norms in Foreign Policy:
Evidence on the Chemical and Nuclear Weapons
“Taboos”*

Christopher W. Blair,[†] Jonathan A. Chu,[‡] and Joshua A. Schwartz[§]

July 18, 2019

Contents

1	Introduction	2
1.1	Pilot Study	2
2	Theory & Hypotheses	2
3	Experimental Design	4
3.1	Assignment to Treatment	6
3.2	Questionnaire	8
4	Sample	14
4.1	Respondent Compensation	14
4.2	Attrition and Attention	14
5	Estimation	15
5.1	Variables	15
5.2	Survey Weights	15
5.3	Treatment Effects	15
5.4	Heterogeneous Effects	16
6	References	16

*Names are in alphabetical order. Equal authorship is implied. Generous support for this research is provided by the Christopher H. Browne Center for International Politics at the University of Pennsylvania.

[†]PhD Candidate, Department of Political Science, University of Pennsylvania, cwblair@sas.upenn.edu

[‡]Postdoctoral Fellow, Perry World House, University of Pennsylvania, jonachu@upenn.edu

[§]PhD Candidate, Department of Political Science, University of Pennsylvania, josha@sas.upenn.edu

1 Introduction

Prominent international relations scholars argue that the use of certain weapons – most notably weapons of mass destruction (WMD) like chemical and nuclear weapons — are taboo (Price 1995; 1997; Tannenwald 1999, 2008). While norms outline standards of appropriate behavior (Finnemore and Sikkink 1998: 891), taboos are stronger than regular norms in that they connote “such qualities as absoluteness, unthinkingness, and taken-for-grantedness” (Tannenwald 1999, 436). However, recent research on the nuclear taboo (Press, Sagan, and Valentino 2013; Sagan and Valentino 2017; Pauly 2018), combined with rampant chemical weapons use in Syria and President Trump’s nuclear saber-rattling against North Korea, brings into question the strength of these supposed “taboos.”

We seek to intervene in this literature by proposing a new conceptualization of norms and then testing this theory in the case of public opinion on the use of chemical and nuclear weapons. In particular, we argue that there may be two types of norm-holders: sincere norm-holders that publicly *and* privately support a norm, and shallow norm-holders that publicly support a norm in order to avoid social sanction, but privately oppose it. Given that the two variants of this norm are observationally equivalent in traditional surveys, existing literature may overestimate the extent to which the public has internalized certain norms. In order to test this argument, we utilize a list experiment design, which is a technique used to estimate people’s opinions about sensitive topics.

1.1 Pilot Study

This pre-analysis plan outlines our design and analyses for a new study that replicates and extends an experiment on public attitudes toward chemical weapons use conducted in March 2016 on Qualtrics. In our proposed study, we have made significant design modifications to the initial study in this plan. First, this design analyzes support for chemical and nuclear weapons, whereas the initial experiment only studied chemical weapons. Second, this design combines list and direct question elicitation for the sensitive items. The previous experiment did not permit simultaneous estimation of the joint survey response to a list experiment and direct question in a multivariate regression context. Third, this design permits us to assess the effect of peer and elite social cues on support for the sensitive item, which the initial study did not.

2 Theory & Hypotheses

Recent research on the nuclear taboo casts doubt on the strength of the general WMD taboo among the public. Most notably, Press, Sagan, and Valentino (2013) argue that public opinion polls that show strong opposition to nuclear use are misleading because they do not force respondents to consider tradeoffs between risking the lives of US soldiers and killing foreign civilians. In a series of survey experiments, they find that large fractions of the American public — and in many cases majorities — would support nuclear use in order to save the lives of US soldiers. For example, 59% of respondents in one scenario would approve of a nuclear strike against Iran that would kill 2 *million* Iranian civilians, but might save the lives of 20,000 American troops (Sagan and Valentino 2017: 58).

We similarly argue that public opinion polls on nuclear and chemical weapons use can be misleading, but for a different reason. Fearing social sanction, some respondents may be unwilling to openly support nuclear or chemical weapons use, even if they privately support it. Consistent with this argument, Pauly (2018: 188-189), in an examination of twenty-six wargames conducted by the US government, finds evidence that one driver of nuclear non-use among elites is reputational costs among peers. For example, in a high-level wargame about Vietnam, a senior observer said, “Well, at the risk of being called a horrible war monger and a few other things . . . I would suggest maybe a small size A-bomb on Dien Bien Phu . . .” (Pauly 2018, 172-173). Besides the speaker recognizing that he has put his reputation on the line, the room responds with laughter to his suggestion, another senior participant implies his proposal is archaic, and the option is rejected (Pauly 2018: 173). Given people’s aversion to censure and embarrassment, we expect that a similar dynamic operates with respect to the public.

In sum, we posit that direct question surveys overestimate public aversion to nuclear and chemical weapons. Therefore, our primary hypothesis – H_1 – is that support for nuclear and chemical weapons will be higher when respondents are asked discreetly rather than directly. In broader terms, this suggests the the existence of two types of norm-holders: sincere norm-holders that publicly *and* privately support a norm, and shallow norm-holders that publicly support a norm, but privately oppose it. This classification scheme could also be applied to other norms where hypocrisy is involved like sovereignty (Krasner 1999) and torture (McKeown 2009).

H_1 : Individuals are more likely to report support for [nuclear/chemical] weapons use when asked indirectly than directly.

We also explore several secondary hypotheses. First, we are interested in investigating whether peer (Kertzer and Zeitzoff 2017) and elite (Zaller 1992; Berinsky 2007) cues shape respondents’ willingness to express support or opposition to nuclear and chemical weapons use. We predict that priming participants with information about peer or elite *support* for allegedly taboo behaviors – like nuclear or chemical weapons use – will increase respondents’ support for those behaviors by lowering their social desirability concerns. This leads to two specific hypotheses. Per H_2 , overall support for nuclear and chemical weapons use should increase when respondents are primed with peer or elite support for these weapons, compared to when respondents are given a generic statement that is unrelated to WMDs. Additionally, per H_3 , we expect that the gap in support for these weapons between the indirect and direct question will decrease when individuals receive the peer or elite cue rather than the unrelated statement. Put more simply, we believe that the peer and elite cues should make respondents that privately support nuclear or chemical weapons more willing to *openly* express their support. Consequently, support for these weapons in the direct question should more closely match support in the indirect question.

H_2 : Individuals that receive the peer or elite cue are more likely to report support for [nuclear/chemical] weapons use than those that receive no cue.

***H*₃: The gap in support for [nuclear/chemical] weapons use between the indirect and direct question will decrease when individuals receive the peer or elite cue rather than no cue.**

Second, we are interested in the impact of several potential moderators: susceptibility to peer influence, age, and partisanship. People that are disposed to care more about what their peers think are more likely to conform to social norms – at least in public – in order to avoid social punishment (Steinberg and Monahan 2007). We thus expect, per *H*₄, that individuals that are more susceptible to peer influence will be less likely to support nuclear or chemical weapons use when asked directly. This hypothesis likely holds because respondents that are more susceptible to peer influence will be more likely to fear the social costs associated with *openly* supporting taboo behavior. Per *H*₅ and *H*₆, we also expect older and Republican respondents to be more likely to support nuclear and chemical weapons for various reasons. For example, older respondents are more likely to have lived through prominent moments when the nuclear or chemical weapons taboos were violated or nearly violated (e.g., Hiroshima and Nagasaki, the Cuban Missile Crisis, and Reagan-era chemical weapons modernization), and Republican administrations have more frequently supported chemical (e.g., the Reagan administration) and nuclear weapons modernization (e.g., the Trump administration).

***H*₄: Individuals that are more susceptible to peer influence will be less likely to support [nuclear/chemical] weapons use when asked directly than individuals that are less susceptible to peer influence.**

***H*₅: Older individuals are more likely to support [nuclear/chemical] weapons use than younger individuals.**

***H*₆: Republicans are more likely to support [nuclear/chemical] weapons use than non-Republicans.**

3 Experimental Design

To test our theory we propose a combined list-direct question survey experimental design (Blair and Imai 2012; Aronow, Coppock, Crawford, and Green 2015; Eady 2017). Specifically, we conduct two list experiments paralleling two direct questions. In the first list-direct question pair we investigate the nuclear weapons taboo (Study A). In the second pair we investigate the chemical weapons taboo (Study B). After being randomized with equal probability into Study A or Study B, respondents in both studies are then randomized into one of three primes. We provide more details about the randomization in Section 3.1. The first prime is the peer cue. Respondents in this group are told: “Many American citizens from both political parties believe that the US should be willing to use [nuclear/chemical] weapons during war.” This cue is meant to lower respondent’s social desirability concerns by informing them that many other Americans believe that [nuclear/chemical] weapons may be justified during war. The second prime is the elite cue. Respondents in this group are told: “Many American government officials from both political parties believe that the US should be willing to use [nuclear/chemical] weapons during war.” This cue is meant to lower respondent’s social desirability concerns by informing them that many elites believe that [nuclear/chemical] weapons

may be justified during war. The third prime is generic, and reads as follows: “The US government does not have an embassy in the country Bhutan.” This prime is related to international relations, but should not affect respondents’ willingness to express support for [nuclear/chemical] weapons use, and does not inform respondents about levels of peer or elite support.

Once respondents are randomized into Study A or Study B and to the peer, elite, or no social prime, we elicit responses to a host of demographic and related pre-test questions, and administer the social prime. In order to administer the social prime we give respondents the following statement: “Below is some information about international politics. You may have heard of some but not others, or none of them at all. Select the ones you’ve heard of before.” The response options are:

- Myanmar (Burma) was once a British colony.
- NATO has 29 member states, which include the US, Canada, and several European countries.
- Peer Cue/Elite Cue/No Cue.

After completing the sociodemographic battery and receiving the social prime, respondents are randomly assigned with equal probability to the treatment or control list. The control group receives the non-sensitive list for its respective study. The treatment group receives the sensitive list for its respective study. Non-sensitive list items are similar to or drawn from non-sensitive list items in the extant literature ([Glynn 2013](#)). Moreover, they are unrelated to nuclear/chemical weapons use, and are relatively low variance ([Li 2019](#)). The control list is presented as follows: “Below are three policies that the US government could adopt. HOW MANY of them do you support? You do not need to specify which ones, just HOW MANY.”

- Lowering taxes on corporations.
- Trading less with other countries.
- Legalizing marijuana for non-medical use.

The sensitive list is identical to the control list, but includes one additional item, the sensitive item corresponding to Study A or Study B respectively. The treatment list reads: “Below are three policies that the US government could adopt. HOW MANY of them do you support? You do not need to specify which ones, just HOW MANY.”

- Lowering taxes on corporations.
- Trading less with other countries.
- Legalizing marijuana for non-medical use.
- Be willing to use [nuclear/chemical] weapons during war.

Following both the control and treatment lists, respondents are administered with an attention check question. Including an attention question ensures respondents are reading and carefully considering the items in the list. Our attention check is drawn from Eady (2017), and asks respondents to identify one of the control items from the list they were shown. Specifically, the attention check reads: “The previous question contained a list of statements. Which of the following subjects was a part of that list?”

- The power of labor unions.
- Gay marriage.
- Legalizing marijuana for non-medical use.
- Don’t know.

After responding to their assigned list and attention check, all respondents are administered a direct question. We opt for a design that combines direct and list experimental responses because these designs enhance efficiency relative to simple difference-in-means (Aronow et. al. 2015), improve confidence in results (Blair, Imai, and Lyall 2014), and enable tests of key design assumptions (Eady 2017). Following Eady’s (2017) recommendation, we include our direct question after the list, rather than randomizing the order of the list and direct questions as in Aronow et. al. (2015). As Eady (2017: 245) explains, by presenting the direct question after the list we can “straightforwardly test whether receiving the treatment list affects the response to the direct question, and if such a relationship exists, to account for this by modeling it in [a] sub-model” tailored to detect misreporting. Also following Eady (2017), we separate the list from the direct question about the sensitive item by several unrelated questions. Designing the survey this way minimizes the extent to which “those in the treatment group recall their previous response to the sensitive item in the list experiment and provide the same response to the direct question for reasons of cognitive ease or to be consistent on principle” (Eady 2017: 245). The direct question reads: “Do you think the US government should adopt the following policy? Policy: "Be willing to use [nuclear/chemical] weapons during war.” The main experimental conditions are summarized in Figure 1.

Following the direct question, all respondents are given an open-ended question that asks them to explain their position on nuclear or chemical weapons. We plan to use this question for supplementary analysis. Finally, respondents *that indicate support* for nuclear or chemical weapons on the direct question will be asked the following behavioral question: “Would you be willing to put your name down on a public petition asking the US government to be willing to use [nuclear/chemical] weapons during war?” Although we do not have strong *a priori* expectations associated with this question, if a significant number of respondents answer “no,” then that would suggest that our design may actually be underestimating the share of the population that are shallow norm-holders. Although respondents may be willing to publicly indicate support for taboo policies when asked, they may be unwilling to take active steps in order to promote those policies due to fear of social punishment.

3.1 Assignment to Treatment

To recap, respondents are first assigned with equal probability to Study A on the nuclear weapons taboo or Study B on the chemical weapons taboo. Next, respondents

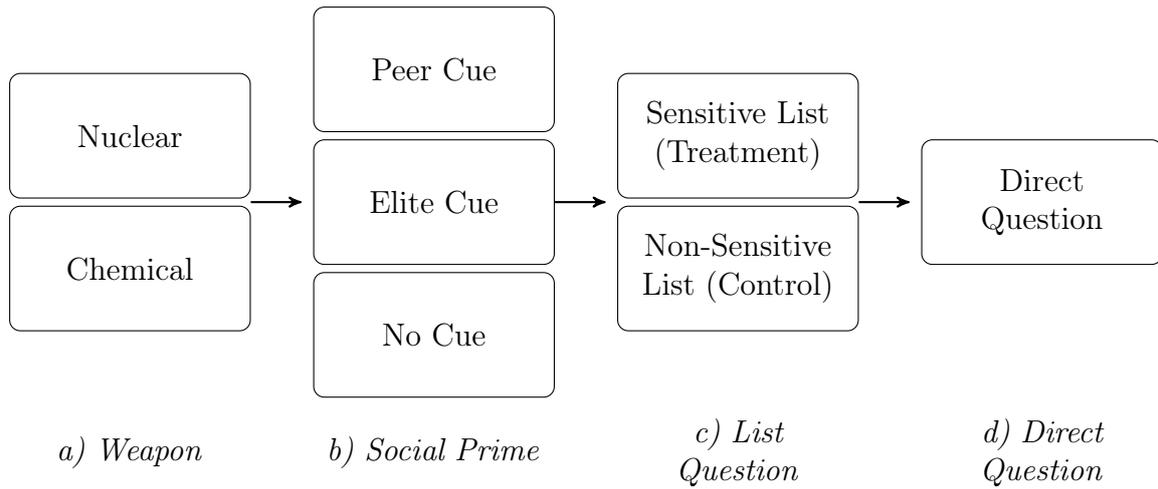


Figure 1: Summary of Experimental Conditions

are assigned to one of the three primes – peer, elite, or none. Respondents are assigned to these primes with probability 0.25, 0.25, and 0.5, respectively. Finally, respondents are assigned to the control or treatment list with equal probability. In [Tables 1 and 2](#) we show the probability of assignment to each experimental condition, along with the expected number of subjects in each cell given an original sample size of about 4,200. The actual number of subjects reported in the tables may vary slightly depending on the survey administration.

Table 1: Probability of Assignment to Each Experimental Condition – Study A

Nuclear Weapons (N = 2100)					
Peer Cue (N = 525)		Elite Cue (N = 525)		No Cue (N = 1050)	
Control (N = 263)	Treatment (N = 262)	Control (N = 263)	Treatment (N = 262)	Control (N = 525)	Treatment (N = 525)
0.0625	0.0625	0.0625	0.0625	0.125	0.125

Table 2: Probability of Assignment to Each Experimental Condition – Study B

Chemical Weapons (N = 2100)					
Peer Cue (N = 525)		Elite Cue (N = 525)		No Cue (N = 1050)	
Control (N = 263)	Treatment (N = 262)	Control (N = 263)	Treatment (N = 262)	Control (N = 525)	Treatment (N = 525)
0.0625	0.0625	0.0625	0.0625	0.125	0.125

3.2 Questionnaire

1. What is your age?

Numerical Entry

2. What is your gender?

Male

Female

Other

3. What is the highest level of education that you have completed?

Less than high school diploma

High school diploma or GED

Some college (no degree)

2-year college degree

4-year college degree

Graduate degree

4. Choose one or more of the races that you consider yourself to be.

White/Caucasian

Black/African American

Latino/Latina

American Indian/Alaska Native

Asian

Native Hawaiian/Pacific Islander

Other

Text Entry

5. What is your total annual income before taxes?

Less than \$20,000

\$20,001-\$45,000

\$45,001-\$70,000

\$70,001-\$100,000

\$100,001-\$200,000

\$200,001 or more

6. Generally speaking, do you think of yourself as a Republican, a Democrat, an Independent, or something else?

Republican

Democrat

Independent

Other

Text Entry

[**Note:** Questions 7 through 14 are in 2 question pairs. The second question in each pair branches from the first. The order of pairs will be randomized.]

7. Please select the statement that best describes you:

Describes Me Better

Some people would do something that they knew was wrong
just to stay on their friends' good side.

BUT

Other people would not do something they knew was wrong
just to stay on their friends' good side.

8. How much does the statement describe you?

Somewhat True of Me Very True of Me

[Last selection]

9. Please select the statement that best describes you:

Describes Me Better

Some people hide their true opinion from their friends if they think their friends will make fun of them because of it.

BUT

Other people will say their true opinion in front of their friends, even if they know their friends will make fun of them because of it.

10. How much does the statement describe you?

Somewhat True of Me Very True of Me

[Last selection]

11. Please select the statement that best describes you:

Describes Me Better

Some people say things they don't really believe because they think it will make their friends respect them more.

BUT

Other people would not say things they didn't really believe just to get their friends to respect them more.

12. How much does the statement describe you?

Somewhat True of Me Very True of Me

[Last selection]

13. Please select the statement that best describes you:

	Describes Me Better
Some people think it's more important to be an individual than to fit in with the crowd.	<input type="radio"/>
BUT	
Other people think it is more important to fit in with the crowd than to stand out as an individual.	<input type="radio"/>

14. How much does the statement describe you?

	Somewhat True of Me	Very True of Me
[Last selection]	<input type="radio"/>	<input type="radio"/>

Social Prime:

15. Below is some information about international politics. You may have heard of some but not others, or none of them at all. Select the ones you've heard of before. [Order of list items randomized.]

- Myanmar (Burma) was once a British colony.
- NATO has 29 member states, which include the US, Canada, and several European countries.
- Peer Cue/Elite Cue/No Cue.

List Experiment:

16. Below are [three/four] policies that the US government could adopt. HOW MANY of them do you support? You do not need to specify which ones, just HOW MANY. [Order of list items randomized; red item only shown to treatment group.]

- Lowering taxes on corporations.
- Trading less with other countries.
- Legalizing marijuana for non-medical use.
- **Be willing to use [nuclear/chemical] weapons during war.**

None of them (0)

One of them (1)

Two of them (2)

Three of them (3)

Four of them (4)

Attention Check:

17. The previous question contained a list of statements. Which of the following subjects was a part of that list?

- The power of labor unions.
- Gay marriage.
- Legalizing marijuana for non-medical use.
- Don't know.

Unrelated Questions:

18. Please indicate your responses to the following questions.

	Never	Once a Month or Less	Once a Week	Every Few Days	Everyday
I drink coffee.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I drink alcohol.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I eat spicy food.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I eat meat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

19. Please indicate your responses to the following questions.

	Definitely Not	Probably Not	Don't Know/Not Sure	Probably Yes	Definitely Yes
I will fly on an airplane in the next year.	<input type="radio"/>				
I will travel abroad in the next year.	<input type="radio"/>				

Direct Question:

20. Do you think the US government should adopt the following policy?

	No	Not Sure	Yes
Be willing to use [nuclear/chemical] weapons during war.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Open-Ended Question:

21. Please write a sentence or two about your response to the last question.

Text Entry

Behavioroid:

22. Would you be willing to put your name down on a public petition asking the US government to be willing to use [nuclear/chemical] weapons during war?" [Only for respondents answering Yes to the direct question.]

- Yes – click petition link.
- No.

4 Sample

In this study, the population of interest is all American adults aged 18 or older. We will field our survey experiment on Amazon’s mTurk platform, and aim for 4,200 respondents. The number of respondents per cell is described in [Table 1](#). All mTurk workers aged 18 or older are eligible to enroll, provided they have a HIT Approval Rate of at least 95%. We implement a HIT Approval Rate threshold for participation to ensure we sample high-quality, attentive workers. Because we will field our survey on mTurk, we have a convenience sample, rather than a nationally representative one. Several studies have found that credible inferences can be drawn from mTurk samples ([Berinsky, Huber, and Lenz 2012](#); [Krupnikov and Levine 2014](#); [Mullinix et al. 2016](#)), but there are still possible concerns about the external validity of our findings given that mTurk subjects differ from the general U.S. population in a number of ways. In general, mTurk skews Democratic, young, and educated ([Berinsky, Huber, and Lenz 2012](#); [Huff and Tingley 2015](#)). However, we believe that the demographics of mTurk actually make this a *harder* test of our hypotheses. As detailed above, we expect that our results should be stronger among older respondents. Hence, a more representative sample would likely reveal even stronger effects than our mTurk sample.

A secondary concern is that because our survey is conducted via the Internet, rather than by phone or in person, differences in the mode of survey administration could bias our results. We do not believe this is a significant concern because, as [Lind et. al. \(2013\)](#) show, online survey takers might be even more susceptible to social desirability. This again makes an mTurk sample a harder test for our theory.

4.1 Respondent Compensation

Our respondents will be compensated with about \$0.50 upon completion of the survey. We anticipate the survey taking 3 to 4 minutes, so this compensation translates to between \$7.50 and \$10.00 per hour. As an hourly rate, this compensation exceeds the minimum wage, and is comparatively high for mTurk studies. Moreover, this compensation level follows that used in recent studies, such as [Aronow, Baron, and Pinson \(2019\)](#).

4.2 Attrition and Attention

Attrition is unlikely to be a significant problem in our survey because it is short – we estimate it will only take 3 to 4 minutes. However, in the event that respondents cease participation prior to completion, or if they fail to answer one or more questions, we will drop them from the sample. This decision follows [Aronow et. al. \(2015: 52\)](#). Likewise, following standard practice in self-administered studies ([Berinsky, Margolis, and Sances 2014](#); [Aronow et. al. 2015: 52](#)), we include an attention check question to mitigate the likelihood that subjects respond without reading or considering the list carefully. We will conduct analyses excluding respondents who attrite from the survey or fail the attention check administered directly following the administration of their assigned list. However, some recent work questions the practice of dropping respondents who fail attentions checks, arguing that doing so induces bias ([Aronow, Baron, and Pinson 2019](#)). Pursuant to this scholarship, we will also conduct analyses without excluding respondents who fail their attention check.

5 Estimation

5.1 Variables

Of the pre-treatment variables, we will treat non-responses and responses of "Don't know" as missing values. Similarly, our four items from the Resistance to Peer Influence scale (Steinberg and Monahan 2007) will be combined into an additive index. For analysis, we will take two approaches to dividing respondents by their score on the scale. First, we will split the scale at its median, defining scores at or above the median as highly resistant to peer influence, and scores below the median as susceptible to peer influence. Second, we will split the scale by the interquartile range, defining the top quarter of scores as highly resistant to peer influence, and the bottom quarter of scores as susceptible to peer influence. Further, our age variable will be split into the following five categories: 18-29, 30-39, 40-49, 50-64, 65+.

5.2 Survey Weights

Our sample is not nationally-representative. As such, we will report results without weights, and, for robustness, with post-stratification weights. Specifically, in specifications with weights we will follow Hainmueller's (2012) approach for entropy balancing against the U.S. Census. We will balance on our core sociodemographic variables: gender, age, race, education, and income status.

5.3 Treatment Effects

A number of different strategies are available for estimating results from list experiments. In the simplest, univariate formulation, two-sample, one-tailed, equal variances t-tests are used to report differences in means between support for the sensitive item in the list experiment versus in the direct question. As a first cut, we too will follow this baseline strategy, comparing all treated respondents to the control, comparing treated respondents to control respondents within each prime, and comparing treated and control respondents across primes.

However, in our primary specifications we will employ the multivariate, maximum likelihood estimator described by Blair and Imai (2012), and the joint estimator described by Aronow et. al. (2015). These are implemented with the functions `ictreg` and `combinedListDirect` from the R package `list`, respectively. When using `combinedListDirect`, we will also report results from the two placebo tests automatically wrapped into the function. Respectively, the placebo tests indicate whether the list estimate is significantly different from 1 among those who respond affirmatively on the direct question, and whether the responses to the direct question are affected by the treatment list. As an additional test, we will report the results from Eady's (2017) maximum likelihood estimator, which allows us to detect whether respondents are misreporting – that is, whether respondents are providing a truthful answer to the list, but failing to reveal their responses when questioned directly. This estimator is available as `listExperiment` from the R package `misreport`.

Finally, for robustness, we will report multivariate estimates from the non-linear least squares and Bayesian estimators described by Blair and Imai (2012), which are

also available in the R package `list`. When using all multivariate estimators described above, we will report results with and without covariate adjustment, opting for covariate adjustment in our primary specifications. We will include the following pre-treatment covariates in covariate adjusted models: age, resistance to peer influence, an indicator for gender, indicators for the categories of race, indicators for the categories of education, indicators for the categories of income status, and indicators for partisan identification.

5.4 Heterogeneous Effects

Per the hypotheses outlined in [Section 2](#), we anticipate heterogeneous effects by responsiveness to peer influence, age, and partisanship. We will explore these heterogeneous effects in the univariate and multivariate tests described in [Section 5.3](#). In large measure, our work is descriptive, and we intend to investigate and display effects by all pre-treatment covariates for which we adjust. However, we do not have strong, *a priori* expectations about heterogeneous effects by gender, race, education, or income status. Our pilot results indicate that women, non-college educated respondents, and respondents with lower incomes are somewhat more supportive of chemical weapons use in a list experimental setting.

6 References

1. Aronow, Peter M., Jonathon Baron, and Lauren Pinson. 2019. "A Note on Dropping Experimental Subjects who Fail a Manipulation Check." *Political Analysis* (Online First).
2. Aronow, Peter M., Alexander Coppock, Forrest W. Crawford, and Donald P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3(1): 43-66.
3. 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.
4. Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3): 351-68.
5. Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3): 739-53.
6. Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1): 47-77.
7. Blair, Graeme, Kosuke Imai, and Jason Lyall. 2014. "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science* 58(4): 1043-63.
8. Eady, Gregory. 2017. "The Statistical Analysis of Misreporting on Sensitive Survey Questions." *Political Analysis* 25(2): 241-59.

9. Finnemore, Martha, and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52(4): 887-917.
10. Glynn, Adam N. 2013. "What Can We Learn With Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77(1): 159-72.
11. Hainmueller, Jens. 2011. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1): 25-46.
12. Huff, Connor, and Dustin Tingley. 2015. "Who are these people? Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3): 1-12.
13. Kertzer, Joshua D., and Thomas Zeitzoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3): 543-558.
14. Krasner, Stephen D. 1990. *Sovereignty: Organized Hypocrisy*. Princeton, NJ: Princeton University Press.
15. Krupnikov, Yanna, and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1(1): 59-80.
16. Li, Yimeng. 2019. "Relaxing the No Liars Assumption in List Experiment Analyses." *Political Analysis* (Online First).
17. Lind, Laura H. Lind, Michael F. Schober, Frederick G. Conrad, and Heidi Reichert. 2013. "Why Do Survey Respondents Disclose More When Computers Ask the Questions?." *Public Opinion Quarterly* 77(4): 888-935.
18. McKeown, Ryder. 2009. "Norm Regress: US Revisionism and the Slow Death of the Torture Norm." *International Relations* 23(1), 5-25.
19. Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2016. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2): 109-38.
20. Pauly, Reid B.C. 2018. "Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint." *International Security* 43(2): 151-92.
21. Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107(1): 188-206.
22. Price, Richard. 1995. "A Genealogy of the Chemical Weapons Taboo." *International Organization* 49(1): 73-103.
23. Price, Richard. 1997. *The Chemical Weapons Taboo*. Ithaca, NY: Cornell University Press.
24. Price, Richard. 2019. "Syria and the Chemical Weapons Taboo." *Journal of Global Security Studies* 4(1): 37-52.

25. Sagan, Scott D., and Benjamin A. Valentino. 2017. "Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants." *International Security* 42(1): 41-97.
26. Steinberg, Laurence, and Kathryn C. Monahan. 2007. "Age Differences in Resistance to Peer Influence." *Developmental Psychology* 43(6): 1531-43.
27. Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3): 433-68.
28. Tannenwald, Nina. 2008. *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945*. New York, NY: Cambridge University Press.
29. Wendt, Alexander. 1999. *Social Theory of International Politics*. New York, NY: Cambridge University Press.
30. Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York, NY: Cambridge University Press.