Abstraction in Experimental Design: Testing the Tradeoffs
Supplementary Appendix

Ryan Brutger[1], Joshua D. Kertzer[2], Jonathan Renshon[3], & Chagai M. Weiss[4]

Last updated: March 3, 2022

[1]Assistant Professor, University of California, Berkeley, Department of Political Science, Email: brutger@berkeley.edu. Web: https://sites.google.com/berkeley.edu/brutger/.

[2]Professor of Government, Department of Government, Harvard University. Email: jkertzer@gov.harvard.edu. Web: http:/people.fas.harvard.edu/~jkertzer/

[3]Associate Professor & Glenn B. and Cleone Orr Hawkins Chair, Department of Political Science, University of Wisconsin-Madison. Email: renshon@wisc.edu. Web: http://jonathanrenshon.net

[4]PhD Candidate, Department of Political Science, University of Wisconsin-Madison, Email cmweiss3@wisc.edu, Web: http://chagaimweiss.com

Contents

List of Tables

List of Figures

Appendix A

Supplementary materials

A.1 Survey overview

The experiments analyzed in our book were embedded in two separate waves (as well as a pre-test wave), implemented in Spring 2019, and Spring 2020. As depicted in Table A.1, we first implemented a pilot study on Amazon Mechanical Turk, to empirically assess different countries' schema and treatment consistency for the Democratic Peace and Nuclear Weapons experiments. Then, in Spring 2019, we implemented a survey on Dynata, including extensions of three experiments: the Nuclear Weapons, In-Group Favoritism, and Democratic Peace experiments. In June 2020, we implemented another survey on Lucid, including extensions of the Democratic Peace and Elite Cue experiments, as well as our three new studies: the Vice President experiment, the Protest experiment, and the WHO experiment. The implementation of our main survey waves followed a shared procedure detailed in Figure A.1.

The platforms we used to collect data are widely used by social scientists. Political scientists have frequently fielded studies using Dynata (formerly known as Survey Sampling International (SSI)) (Kam, 2012; Malhotra, Margalit and Mo, 2013; Kertzer and Brutger, 2016; Brutger and Rathbun, 2021) and Lucid (Tomz and Weeks, 2020; Hill and Huber, 2019; Orr and Huber, 2020), to implement vignette based survey

Table A.1: Description of experiments by survey wave

	Pre-test	***Wave 1***	***Wave 2***
Test of Country Names	x		
Nuclear Weapons		x	
In-Group Favoritism		x	
Democratic Peace		x	x
Elite Cues			x
Vice President			x
Protest			x
WHO			x

1. **Informed consent and screening:** Respondents are asked to consent to the study, and are screened out if they are located outside of the US. In wave 1 we further screened respondents using a mobile device to answer the survey.

2. **Assignment to situational hypotheticality treatment:** Respondents are assigned to either an explicitly or implicitly hypothetical condition in our first wave. In our second wave we randomized whether scenarios were described as explicitly hypothetical, real, or a pure control condition where no description of hypotheticality was mentioned in the introduction the experiment. This treatment was varied across respondents but remains constant across all studies for a given respondent. To strengthen this treatment, the emphasis on hypotheticality recurs in follow up questions that mention the experimental scenario.

3. **Assignment to order of experiments:** In both studies we randomized the order of studies to avoid ordering effects.

4. **Assignment to original study-level treatments:** Respondents were randomly assigned to the original conditions of studies. Unlike the assignment of the hypotheticality treatment, this assignment is independent across all studies.

5. **Assignment to contextual detail/actor identity treatments:** Respondents were randomly assigned versions of the original/new studies that vary in their amount of contextual detail and/or in the identities of the actors in the scenarios. Unlike the situational hypotheticality treatment, this assignment is independent across all studies.

6. **Pre-Treatment Covariate Collection:** Respondents answered a battery of pre-treatment covariates. In study 2, these include affect towards a host of countries.

7. **Experiment completion:** Respondents participated in experiments and responded to our main outcome measures detailed below. Outcomes include original survey items as well as additional questions which investigated respondents' attention to the general vignette context and treatment.

8. **Additional Demographic and individual difference batteries:** In wave 1, respondents respond to covariate batteries relating to: Foreign policy attitudes, cooperative internationalism, need for cognition, cognitive reflection.(Thomson and Oppenheimer, 2016), political knowledge (Clifford and Jerit, 2016), and demographics. In wave 2, respondents respond to pretreatment questions gathering information about political knowledge and feelings toward different actors, and post-treatment questions gathering additional demographics.

Figure A.1: Overview of Study Protocol

experiments. Furthermore, recent investigations suggest that Lucid is a suitable platform for implementing survey experiments in the U.S. context (Coppock and McClel-

lan, 2019), and have found that experiments fielded on Lucid before the COVID-19 pandemic replicated during the COVID-19 pandemic as well (Peyton, Huber and Coppock, 2021). While the sampling strategies of both survey providers differ from one another, they both involve online panels. We report descriptive statistics for both our samples in Tables A.2-A.3.

Table A.2: Descriptive Statistics - Wave I

Statistic	N	Mean	St. Dev.	Min	Max
Age	4,289	51.040	17.052	18.000	99.000
Male	4,330	0.469	0.499	0.000	1.000
Female	4,330	0.525	0.499	0.000	1.000
Education	4,317	3.645	1.650	1.000	8.000
White	4,320	0.797	0.403	0.000	1.000
Black	4,320	0.082	0.274	0.000	1.000
Hispanic	4,320	0.043	0.203	0.000	1.000
Asian	4,320	0.050	0.218	0.000	1.000
Democrat	4,330	0.343	0.475	0.000	1.000
Republican	4,330	0.305	0.461	0.000	1.000
Independent	4,330	0.274	0.446	0.000	1.000

Table A.3: Descriptive Statistics - Wave II

Statistic	N	Mean	St. Dev.	Min	Max
Age	4,025	45.236	17.262	18.000	98.000
Male	4,026	0.474	0.499	0.000	1.000
Female	4,026	0.517	0.500	0.000	1.000
Education	3,997	4.588	1.945	1.000	8.000
White	4,028	0.724	0.447	0.000	1.000
Black	4,028	0.117	0.321	0.000	1.000
Hispanic	4,028	0.072	0.259	0.000	1.000
Asian	4,028	0.042	0.201	0.000	1.000
Democrat	4,026	0.349	0.477	0.000	1.000
Republican	4,026	0.343	0.475	0.000	1.000
Independent	4,026	0.233	0.423	0.000	1.000

A.2 Study instrumentation – extension studies

In this section we describe the main experimental vignettes that we employ for our four main replication and extension studies reported in the main text.

A.2.1 Nuclear Weapons experiment

The NUCLEAR WEAPONS experiment extends Press, Sagan and Valentino's (2013) examination of norms against the use of nuclear weapons in public opinion, to study the effects of both actor identity and contextual detail in tandem. The original study investigated whether normative prohibitions against the use of nuclear weapons were a factor in the U.S. public's preferences about whether and how to use force in world politics. It did so by randomizing the relative probability of success for conventional attacks relative to nuclear attacks. The original study includes three treatment conditions where conventional attacks have a 90%, 70% or 45% success probability in relation to nuclear attacks which are held constant at 90% probability of success. The original study also randomizes whether the scenario is described as prospective or retrospective. For the sake of simplicity we only consider the 90% and 45% success probability conditions, and we described scenarios as prospective (randomizing whether it is portrayed as explicitly or implicitly hypothetical).

In our extension of the NUCLEAR WEAPONS experiment, we consider the consequences of situational hypotheticality, contextual detail, and actor identity, adding three additional treatment arms to the original study. Apart from our hypotheticality treatment which is randomized at the respondent level and consistent throughout all vignettes in wave 1, we manipulate the vignette's context to either include: (1) elaborate context (as in the original study) or (2) reduced context. We also consider four alternative country names, which include: (1) Syria (as in the original study)—a schema consistent country, (2) an unnamed country ("a foreign country"), (3) a fictitious country name ("Malaguay"), or (4) a real and schema-inconsistent country (Bolivia). Thus, we present all subjects with the following text below, as well as a

summary table (see Table A.4). For ease of interpretation, we depict the original study treatment in **bold red**, the hypotheticality treatment in green, the context level treatment in *italic blue*, and the country level treatment in blue.

There is much concern these days about the spread of nuclear weapons. We are going to describe a hypothetical situation / situation the United States 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.

Joint Chiefs Report Concludes Nuclear and Conventional Options for Destroying Al Qaeda Nuke Lab Equally Effective / Joint Chiefs Say U.S. Nuclear Options Offers Dramatically Increased Chances of Destroying Nuke Lab

Expected Civilian Casualties, Physical Destruction Equivalent for Both Options / Chiefs Conclude Nuclear Option Has 90% Chance of Success, Conventional Only 45%

The Associated Press

A report from *General Martin Dempsey, Chairman of the Joint Chiefs of Staff, / the Joint Chiefs of Staff* to the President **concludes that military strikes using nuclear or conventional weapons would be "equally effective" / concludes that nuclear weapons would be "dramatically more effective" than conventional strikes** in destroying an Al Qaeda nuclear weapons facility in Syria / Malaguay / the country / Ecuador.

The report compares two American military options, a conventional strike using nearly one hundred conventionally-armed cruise missiles, and an attack using two small, nuclear-armed cruise missiles. **The report estimates that both options have a 90 percent chance of successfully destroying the Al Qaeda nuclear weapons lab / the conventional strike has a 45 percent chance of successfully destroying the atomic bomb lab while nuclear weapons increase the chances of success to approximately 90 percent**. Empty / Syria / Malaguay / the country / Ecuador has refused to allow international inspectors access to the facility.

The Joint Chief's assessment comes two weeks after Russian intelligence agents intercepted a shipment of centrifuges and low-enriched uranium which could be used to produce nuclear weapons. The bomb-making equipment was being smuggled out of Russia to an Al Qaeda facility located near a remote town in the north of Syria / Malaguay / the country / Ecuador. *The suspects in the smuggling operation were employed at a Russian nuclear lab. The smugglers confirmed under questioning that other shipments of centrifuges and low-enriched uranium had already been delivered to the Al Qaeda base, where the centrifuges are being used to make fuel for a nuclear bomb. The smugglers stated that there will be enough bomb grade material produced for at least one weapon within two weeks.* Syria / Malaguay / the country / Ecuador *has refused to allow international inspectors access to the facility./ Empty*

The Joint Chiefs' report to the President does not recommend a specific course of action, *However, it concludes that "because the Al Qaeda facility is comprised of a series of deeply*

buried bunkers, a strike would require either large numbers of conventional missiles, or two nuclear weapons, to destroy the facility." / but concludes that destroying the facility would require either large numbers of conventional missiles, or two nuclear weapons.

Either option would have roughly a ninety percent chance of success, according to the report. / According to the report, because of the facility's depth, nuclear weapons would be far more effective for destroying the target.

The report was leaked to the Associated Press by a high-ranking administration official involved in planning the strike. According to the official, the centrifuges and nuclear materials are too large to be moved without detection. / Empty The US intelligence official stated that he has high confidence that Al Qaeda is within two weeks of producing an operational bomb. *After that, the official said, "all bets are off." According to Dr. David Wright, a nuclear expert at the Union of Concerned Scientists, an independent think-tank based in Washington, D.C., "If a bomb of this size exploded in New York City, it could easily kill 50,000 to 70,000 people." / ; estimates suggest that if a bomb of this size exploded in New York City, it could easily kill 50,000 to 70,000 people.*

The report states that the remote location of the Al Qaeda facility should limit civilian fatalities in Syria / Malaguay / the country / Ecuador *for either option. Because many conventional weapons would be required to destroy the Al Qaeda base, the report estimates that "the two options would kill approximately the same number of* Syrian / Malaguayian / foreign / Ecuadorian *civilians" ; about 1,000, including immediate deaths and long term consequences of the conventional and nuclear strike. As both options will rely on cruise missiles launched from U.S. naval vessels, the report concludes that "no U.S. military personnel are at risk in either operation." / The report states that* Syrian / Malaguayan / the country's / Ecuadorian *civilian fatalities would be limited to about 1,000 for either option, including immediate deaths and long term consequences of the conventional and nuclear strike. No U.S. military personnel would be at risk in either operation.*

Table A.4: Table accompanying Nuclear Weapons experiment

Target: Al Qaeda Nuclear Weapons		
	U.S Nuclear Strike	**U.S Conventional Strike**
Probability of Success	90%	**90% / 45%**
Estimated Syrian / Malaguayian / Foreign / Ecuadorian Civilian Deaths	1,000	1,000
IF U.S. STRIKE FAILS 50,000 - 70,000 US. CIVILIAN FATALITIES		
Chart from Joint Chief's report describing nuclear and conventional options for strike on Al Qaeda nuclear lab		

After reading the scenario, respondents reported responses to three outcome questions (see Figure A.2). Our main outcome of interest is preference for a nuclear weapon attack (the last item in Figure A.2).

- Given the facts described in the article, if the United States decided to conduct a nuclear strike to destroy the Al Qaeda base, how much would you approve or disapprove of the U.S. military action? (6 point approve disapprove scale)
- Given the facts described in this article, if the United States decided to conduct a conventional strike to destroy the Al Qaeda base, how much would you approve or disapprove of the U.S. military action? (6 point approve disapprove scale)
- If you had to choose between one of the two U.S. military options described in the article, would you prefer the nuclear strike or the conventional strike?
  - strongly prefer the conventional strike;
  - somewhat prefer the conventional strike;
  - somewhat prefer the nuclear strike;
  - strongly prefer the nuclear strike.

Figure A.2: Nuclear Weapons outcome questions

We also included a follow-up question from the original instrument, which was administered to respondents who indicated a preference for conventional attacks, soliciting their reasons behind their decision. We also included a thought listing exercise (employed after all vignettes), relating to the nuclear weapons vignette (see Figure A.3). Lastly, we asked respondents a manipulation check question (see Figure A.4).

When you think about the scenario/ hypothetical scenario you just read, what features of the scenario/ hypothetical scenario come to mind? Please list these thoughts or considerations below.

Simply write down the first thought that comes to mind in the first box, the second in the second box, and so on. Please put only one idea or thought in a box.

We've deliberately provided more boxes below than we think most people will need, just so you have plenty of room.

Figure A.3: Thought listing question

Think back to the scenario described to you earlier in the survey. What is the relation between the probability of success for nuclear and conventional attacks?
possible responses include:

- Nuclear attacks will be more successful than conventional attacks
- Conventional attacks will be more successful than nuclear attacks
- Conventional and nuclear attacks have similar probabilities of success

Figure A.4: Nuclear Weapons manipulation check

A.2.2 In-Group Favoritism experiment

The In-Group favoritism experiment extends portions of Mutz and Kim's (2017) investigation of American trade preferences, to study the effects of additional contextual detail and situational hypotheticality. In extending their basic framework, we focus on a common decision experimenters grapple with when designing instruments: how much contextual detail should vignettes include? We do so by randomly assigning respondents to either the original short vignette, or a more elaborate vignette which provides further detail on the experimental scenario. Consistent with Bansak et al. (2021), we provide two types of additional context. The first is "filler" context, with peripheral information that increases the volume of text respondents are presented with, but is not expected to interact with the treatment. The second is "charged" context that similarly increases the length of the stimulus, but which is more relevant to the treatment. In so doing, we test how additional information that is either likely or unlikely to interact with the study's main treatment moderates the original findings.

Accordingly, in our In-Group favoritism vignette, we manipulate three different treatments. First, we manipulate whether the scenario is described as implicitly or explicitly hypohtetical. Second, we manipulate a modified version of the original treatment, varying the number of jobs gained or lost by the U.S and an adversary (Possible combinations include: US gains 1,000, other gains 10; US gains 10, other gains 1,000; US gains 10, other loses 1000). Finally, we manipulate the context in the experimental vignette to include either: (1) no additional context, (2) filler context

which is *unlikely* to interact with treatment, or (3) charged context which is *likely* to interact with treatment.

We presented all subjects with the following introduction, along with a vignette whose contents randomly varied across respondents. For ease of interpretation, we depict our hypotheticality treatment in underlined green, the original study level treatment in **bold red**, and the context level treatment in *italic blue*.

There is much concern these days about intentional trade and job security. We are going to describe a hypothetical situation / situation the United States 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. Here is the hypothetical situation / situation: The United States is considering a trade policy that would have the following effects:
For each **1,000 people in the U.S. who gain a job and can now provide for their family, 10 people in a country that we trade with will gain new jobs and now be able to provide for their family / 10 people in the U.S. who gain a job and can now provide for their family, 1,000 people in a country that we trade with will gain new jobs and now be able to provide for their family / 10 people in the U.S. who gain a job and can now provide for their family, 1,000 people in a country that we trade with will lose their jobs and will no longer be able to provide for their family**.[a].[b]
Additional context:
None
Filler Context: If approved, this policy will be implemented within the next two years. As part of the implementation process, a commission of government officials and bureaucrats will outline the financial implications of the policy and provide guidance to businesses on how the new agreement affects them. Lastly, a team comprised of bureaucrats from both countries will oversee the policy implementation process which is expected to last two years. Over the past 20 years, the trade volume between the United States and this country has been steadily increasing. There have been some years where the volume of trade has increased rapidly, while other years it has been somewhat slower. Throughout the past 20 years, both countries have signed several agreements, which were implemented in good faith. Both countries export and import a wide range of products, which will be covered by the terms of the new agreement if it is approved.
Charged Context: If approved, this policy will be implemented within the next two years. Analysis of the agreement has determined that it will dramatically increase trade between the countries. This has the potential to create new business opportunities in both countries, but may also make it harder for some companies to compete. Lastly, a team comprised of bureaucrats from both countries will oversee the policy implementation process which is expected to last two years. Over the past 20 years, the trade volume between the United States and this country has been steadily increasing. More specifically, U.S. goods and service trade with this country totaled an estimated $258.7 billion in 2018. Exports were $121 billion; imports were $137.7 billion. The U.S. goods and services trade deficit with the country was $47.5 billion in 2018. Throughout the past 20 years, both countries have signed several agreements, which were implemented in good faith.

[a]Possible combinations are: US gains 1,000, other gains 10; US gains 10, other gains 1,000; US gains 10, other loses 1000.

[b]Possible combinations are: US gains 1,000, other gains 10; US gains 10, other gains 1,000; US gains 10, other loses 1000.

Figure A.5: In-Group Favoritism vignette

After reading the vignette described in Figure A.5, respondents were exposed to a two-stage outcome measure reported in Figure A.6:

- Would you be likely to support this trade policy or oppose this trade policy? (Support / Oppose)
  - **If support**: Are you strongly supportive of this new trade policy or somewhat supportive of this new trade policy? (Strongly supportive / somewhat supportive)
  - **If oppose**: Are you strongly opposed of this new trade policy or somewhat opposed of this new trade policy? (Strongly opposed / somewhat opposed)

Figure A.6: In-Group Favoritism outcomes

Based on the questions reported in Figure A.6, we created our main DV, measuring support for the described policy, on a four point scale ranging from strongly oppose to strongly support. After collecting our main outcome variable we further asked respondents to engage in a thought listing task. The thought listing task is similar to the one reported in Figure A.3. Following the thought listing exercise detailed above, we directly investigated respondents' attention to their main treatment condition. To do so, we asked the following manipulation check reported in Figure A.7:

Think back to the trade policy that was described to you earlier in the survey. Will our trading partner benefit more than the US, will the US benefit more than the trading partner, or will they be impacted equally?
possible responses include:

- The trading partner will benefit more than the US
- The US will benefit more than trading
- Both countries will benefit equally

Figure A.7: In-Group Favoritism manipulation check

A.2.3 Democratic Peace

In our Democratic Peace extension we manipulated the adversary country name, in additional to the original treatment, and situational hypotheticality. Though the original study manipulated alliance and trade, we fixed those conditions to be constant. We decided to do so for purposes of tractability, since our main objective

in this extension was to consider the consequences of different experimental design choices (varying actor identity).

We fielded two versions of the DEMOCRATIC PEACE extension, one in wave 1 and the other in wave 2. The key difference between waves, was the countries employed in our vignette at each wave. In the vignette of our first wave, we described the adversary country as either unnamed (the country), made up (Malaguay), real schema consistent but treatment inconsistent (Iran), or real schema inconsistent but treatment consistent (Ecuador). In the vignette embedded in the second wave, we described the adversary country as either unnamed (the country), real schema consistent but treatment inconsistent (Iran), or real schema consistent and treatment consistent (Turkey). We present the vignette in Figure A.8. For ease of reading, we denote hypotheticality underlined green, the original democracy treatment in **bold red**, and country name in *Italic Blue*.

The main DV for this study was "Would you favor or oppose using the U.S. military to attack *The country's / Malaguay's / Iran's / Ecuador's / Turkey's* nuclear development sites?" Response options ranged from 1 (oppose strongly) to 5 (favor strongly). We also collected question regarding potential mechanisms originally employed by Tomz and Weeks (2013). After collecting these variables, we asked respondents to participate in a thought listing exercise (as reported in Figure A.3), and implemented a final manipulation check reported in Figure A.9.

A.2.4 Elite Cues experiment

In our ELITE CUES experiment we extent Nicholson's (2012) study of elite cues which focused on immigration reform in the United States. We designed this additional study, to explore the effects of actor identity in experimental design. Nicholson's original study includes several experiments, considering different policies and cue-givers. However, we focus on the immigration policy experiment endorsed by politicians (rather than parties).

Nicholson's original study examined the effect of in/out party endorsements on

There is much concern these days about the spread of nuclear weapons. We are going to describe a situation / real situation / hypothetical situation the United States 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.

Here is the hypothetical situation / situation:

- *A country / Malaguay / Iran / Ecuador/ Turkey* 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;
- *The country / Malaguay / Iran / Ecuador/ Turkey* has not signed a military alliance with the United States;
- *The country /Malaguay / Iran / Ecuador/ Turkey* has high levels of trade with the United States;
- *The country / Malaguay / Iran / Ecuador / Turkey* **is a democracy and shows every sign that it will remain a democracy / is not a democracy and shows no sign of becoming a democracy**;
- *The country's / Malaguay's / Iran's / Ecuador's / Turkey's* nonnuclear military forces are half as strong as U.S. nonnuclear forces;
- *The country's / Malaguay's / Iran's / Ecuador's / Turkey's* motives remain unclear, but if it builds nuclear weapons, it will have the power to blackmail or destroy other countries;
- *The country / Malaguay / Iran / Ecuador / Turkey* has refused all requests to stop its nuclear weapons program.

By attacking the *The country's / Malaguay's / Iran's / Ecuador's/ Turkey's* nuclear development sites now, the United States could prevent the country from making any nuclear weapons.

Figure A.8: Democratic Peace vignette

Think back to the scenario described to you earlier in the survey. Is the country with which the US is engaging a democracy or not?
possible responses include:

- Democracy
- Not a democracy

Figure A.9: Democratic Peace manipulation check

partisan opinion in the context of a proposal to reform U.S. immigration policy that centered on a "path to citizenship" and used high-salience real actors: Barack Obama or John McCain. In our extension, we updated the relevant salient cuegivers (Joe Biden or Donald Trump), while also adding additional actor identity treatments that vary whether the immigration reform endorsement is made by less salient partisan cuegivers (Senator Tom Carper of Delaware or Senator Mike Rounds of South Dakota),

or by a fictional politician (Stephen Smith) whose partisanship we manipulate. We also updated the substantive context of the experiment to focus on protection for "Dreamers" in the U.S.

As part of our experimental vignette, we randomized three treatments: situational hypotheticality (whether the scenario was described as implicitly hypothetical, explicitly hypothetical, or real), original cue treatment (whether the policy was endorsed by an in-party or out-party), and actor identity (whether the endorsing actor was a high salience politician, low salience politician, or a made up politician). We depict our vignette in Figure A.10. For ease of reading, we depict the hypotheticality treatment in underlined green, and our actor treatment (which combines both the partisanship and identity of an actor) in *italic blue*.

There is much concern about immigration policy in American Politics. We are going to describe a situation / real situation / hypothetical situation.
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.
As you know, there has been a lot of talk about immigration reform policy in the news. One proposal *Empty / , backed by Democrat Joe Biden / , backed by Republican Donald Trump / , backed by Republican Mike Rounds / , backed by Democrat Tom Carper / , backed by Democrat Stephen Smith / , backed by Republican Stephen Smith* provided protections for Dreamers-including legal status and a path to legal citizenship for some of them.
What is your view of this immigration policy?" (5 point scale, ranging from strongly support to strongly oppose with higher values associated with opposition)

Figure A.10: Elite Cue vignette

As indicated in Figure A.10, like in the original study, the vignette was followed by our outcome measure, asking respondents to state their level of support for the proposed policy on a five point scale. After collecting the main outcome variable, we asked respondents to complete another thought listing task (See Figure A.3). Then, we turn to directly investigate respondents' attention to their main treatment condition, by employing a factual manipulation check, presented in Figure A.11:

Think back to the **most recent** scenario described to you earlier in the survey. Was the immigration policy described, endorsed by a member of the Democratic party, the Republican party, an independent candidate, or no one at all.

- Endorsed by a member of the Democratic party
- Endorsed by a member of the Republican party
- Endorsed by an independent candidate
- Not endorsed by anyone

Figure A.11: Elite Cue Treatment manipulation check

A.3 Study instrumentation – original studies

Alongside our extensions described above, in wave 2 of our surveys we implemented three additional studies which focused on current events of Spring 2020. Our main objective was to further explore how manipulating situational hypotheticality affects the estimates of the ATEs in each study.

A.3.1 Vice President experiment

In our first original study, we focus on the then Democratic candidate for president, Joe Biden, and his expected vice presidential candidate selection. In doing so, we explore how citizens would respond to the selection of an African American (rather than white) female running mate. To do so, we exposed respondents to a simple vignette, randomizing the racial identity of Biden's expected running mate (white or African American), as well as the framing of the scenario as a hypothetical, current, or future scenario. We report our vignette in Figure A.12, and present our hypotheticality treatment in underlined green, and the main treatment in **bold red**.

After reading the vignette, and responding to our main outcome question (reported in the bottom of Figure A.12), we asked respondents to participate in our thought listing exercise (See Figure A.3). Then, respondents were asked a manipulation check question regarding the gender and racial identity of the prospective vice presidential candidate. We concluded this experiment with a final question, ask-

Next, we ask you to consider that there is much concern regarding elections in the U.S. We will now present you with a scenario relating to elections.
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.
In a hypothetical presidential race, the democratic party candidate has selected / As part of the upcoming 2020 presidential race, leading political experts expect that the democratic candidate Joe Biden will select / In the 2028 presidential race, the democratic party candidate has selected, **a white / an African American** woman as his running mate and future vice president.

How likely are you to vote for this candidate? (4 point likely unlikely scale)

Figure A.12: Vice President vignette

ing respondents if they had a specific politician in mind when reading the vignette; respondents who answered "Yes" to this question were asked to share the specific politician about which they thought while reading the vignette.

A.3.2 Protest experiment

We designed our second original experiment, in the context of the 2020 Black Lives Matter protests, to examine whether respondents' support for a racial justice protest is affected by protesters' use of violent or non-violent action. To that end, in our experiment we randomized how protesters' actions were described (violent or non-violent), as well as the degree of situational hypotheticality in the scenario. Specifically, we frame the scenario as either: hypothetical, real and with experts' prediction of how the protest will evolve, real with a description of current dynamics of the protest, or set in the future. We report our vignette in Figure A.13, where we present our hypotheticality treatment in underlined green, and the main treatment in **bold red**.

After reading about the protest, respondents were asked about their support for the protest's cause (see bottom of Figure A.13). Then, we solicited respondents' reaction to the vignette through a thought listing exercise (See Figure A.3). Finally, respondents were asked a manipulation check question regarding the protesters' behavior.

Next, we ask you to consider that there is much concern regarding protests in the U.S. We will now present you with a scenario relating to protests.

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 social issue.

Please Consider the following hypothetical scenario: Thousands of People across the U.S. have participated in protests, demanding racial justice and condemning police brutality. As part of the demonstrations, protesters are engaging in / Please Consider the following scenario: In recent days, thousands of people across the U.S. have participated in protests, demanding racial justice and condemning police brutality. As part of the demonstrations, political experts and commentators expect that protesters will engage in / Please Consider the following scenario: In recent days, thousands of people across the U.S. have participated in protests, demanding racial justice and condemning police brutality. As part of the demonstrations, protesters are engaging in / Please consider the following scenario which may happen in the future: It is the year 2030 and thousands of people across the U.S. have participated in protests, demanding racial justice and condemning police brutality. As part of the demonstrations, protesters are engaging in, **vandalism, looting, and violent action / peaceful demonstrations which avoid vandalism and looting**.

To what extent do you agree with the following statement: I support the cause of the protests which are demanding racial justice and an end to police brutality. (5 point agree disagree scale)

Figure A.13: Protest vignette

A.3.3 WHO Experiment

Our final original experiment was inspired by ongoing uncertainty relating to the U.S. involvement with the Wold Health Organization (WHO), and Americans' attitudes towards the WHO. As part of the experiment, we randomized whether the U.S. is described as withdrawing or remaining in the WHO, as well as whether the scenario was described as hypothetical, real, or a future scenario. Respondents were exposed to the vignette described in Figure A.14, in which we depict the hypotheticality treatment in underlined green, and the main treatment in **bold red**.

After reading the vignette, and learning about the U.S. engagement with the WHO, respondents were asked about their trust in the WHO, and their belief about the value of the U.S. remaining a WHO member (see bottom of Figure A.14). Then, we solicited respondents' reaction to the vignette through our thought listing exercise (See Figure A.3). Finally, respondents were administered a manipulation check question regarding the U.S. decision to remain or withdraw from the WHO.

Next, we ask you to consider that there is much concern regarding U.S. involvement in international politics. We will now present you with a scenario relating to the U.S.
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 social issue.

Please Consider the following scenario / hypothetical scenario/ scenario which may happen in the future:

There / Over the past several months there / It is the year 2030, and there is much discussion about levels of United States' support for the World Health Organization (WHO) and the United States' continued role in the organization. The U.S. has decided to **withdraw from / remain in**. the World Health Organization

To what extent do you agree with the following statement:
I trust the World Health Organization to do what is right (5 point agree disagree scale)
Is being a member of the World Health Organization good or bad for the U.S.? (5 point good bad scale)

Figure A.14: WHO vignette

A.4 Need for cognition additional results

In this section, we report results from additional models considering the implications of design choices, for subsamples of respondents with low (below average) and high (above average) levels of need for cognition.

In Figure A.15, we consider the moderating effects of country name in our Democratic Peace and Nuclear Weapons experiments, across levels of need for cognition. We find few differences in patterns across our subsamples of respondents with low and high levels of need for cognition, with one exception. That is, in the Democratic Peace experiment, Iran as a country name appears to have a negative and statistically significant moderating effect for the sample of respondents with low levels of need for cognition. While the Iran moderating effect amongst respondents with high levels of need for cognition is still negative, it is smaller in magnitude and imprecisely estimated. We interpret this pattern as suggestive evidence that respondents with low levels of need for cognition are likely driving the main treatment inconsistency effect which we identify in the main text.

In Figure A.16 we examine the moderating effect of situational hypotheticality across our subsamples of respondents with low and high levels of need for cognition.

Figure A.15: Moderating effects of country name by need for cognition

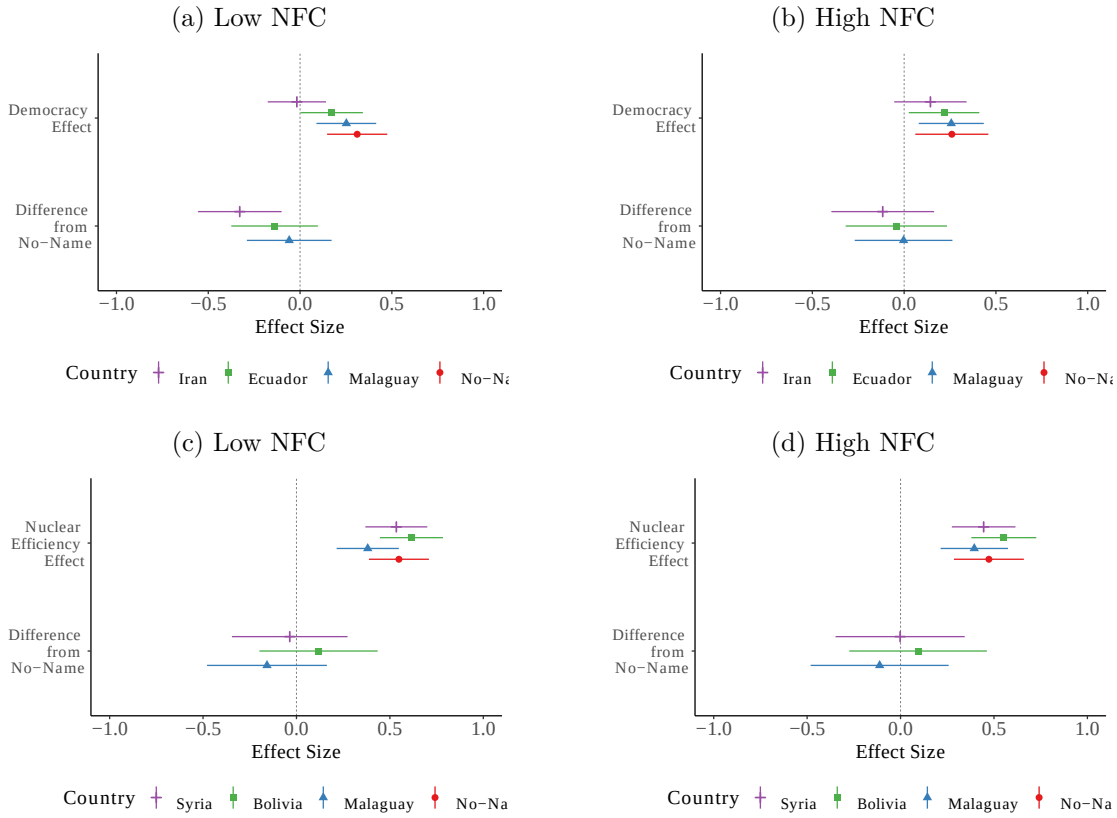

Figure A.15 shows the moderating effects of country name, across different levels of survey respondents' need for cognition in the Democratic Peace (in the top row) and Nuclear Weapons (in the bottom row) experiments. All outcomes are standardized.

For the most part, we find little evidence that variation in hypotheticality moderates average treatment effects, though it appears that framing the In-Group Favoritism experiment as explicitly hypothetical has a small, positive, and precisely estimated moderating effect among those with low need for cognition.

A.5 Political knowledge additional results

In this section, we consider how contextual detail and situational hypotheticality moderate original average treatment effects for subjects with low and high levels of political knowledge. To do so, we subset our sample, and focus on respondents with

Figure A.16: Moderating effects of situational hypotheticality by need for cognition

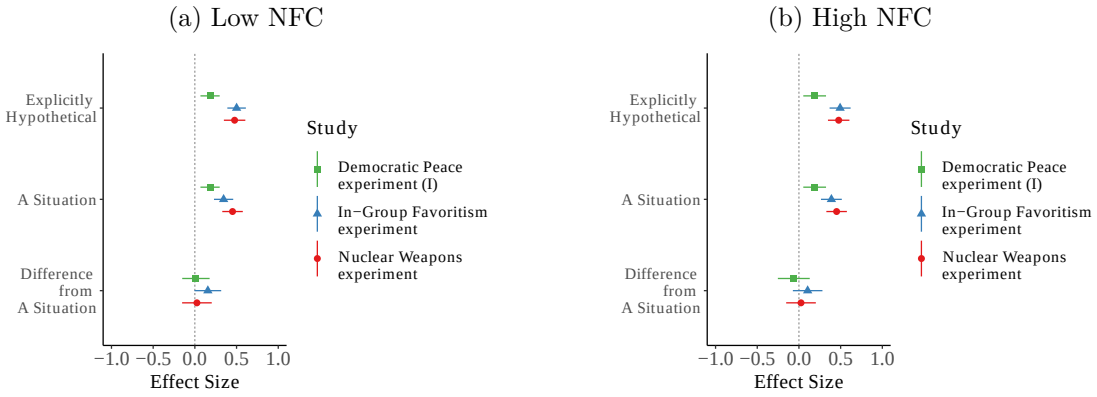

Figure A.16 shows the moderating effects of framing a scenario as explicitly (rather than implicitly) hypothetical, across different levels of survey respondents' need for cognition. All outcomes are standardized.

low or high levels of political knowledge. In Figure A.17, we report results from models which evaluate the moderating effects of context in the In-Group Favoritism and Nuclear Weapons experiments, amongst the samples of respondents with high and low political knowledge. Given the small sample size, some care should be taken in interpreting the results, but the plot shows that the moderating effect of additional context is negatively signed for both high and low-knowledge respondents. Similarly, for the Nuclear Weapons experiment the moderating effect of additional context is negatively signed for both high and low knowledge respondents. Here, the moderating effect attains statistical significance among low knowledge respondents, but the point estimates are similar in both instances. We therefore obtain relatively similar findings for the moderating effects of contextual detail between respondents with high and low levels of political knowledge.

In Figures A.18-A.19, we consider the extent to which framing a given scenario as explicitly rather than implicitly hypothetical (Figure A.18), or real rather than explicitly hypothetical (Figure A.19) moderates average treatment effects in our studies, for respondents with varying level of political knowledge. Consistent with our main results reported in Chapter 4, we do not find evidence that situational hypothetically

Figure A.17: Moderating effects of context detail by political knowledge

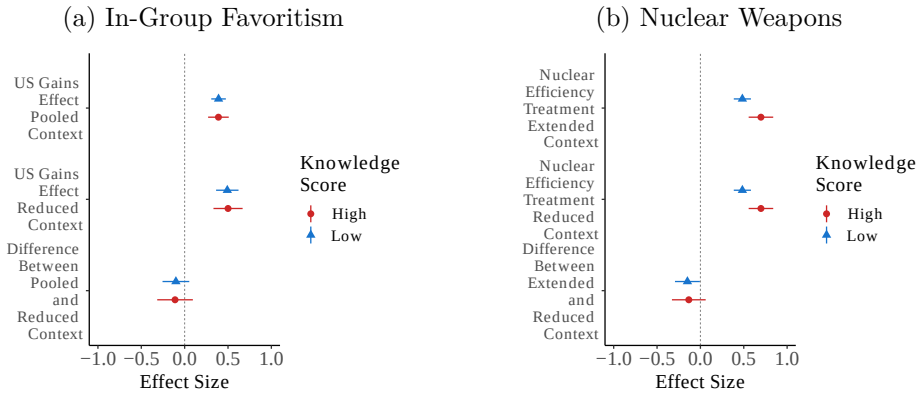

Figure A.17 shows the moderating effects of context, across different levels of survey respondents' political knowledge, in the In-Group Favoritism and Nuclear Weapons experiments. All outcomes are standardized.

Figure A.18: Moderating effects of situational hypotheticality by political knowledge

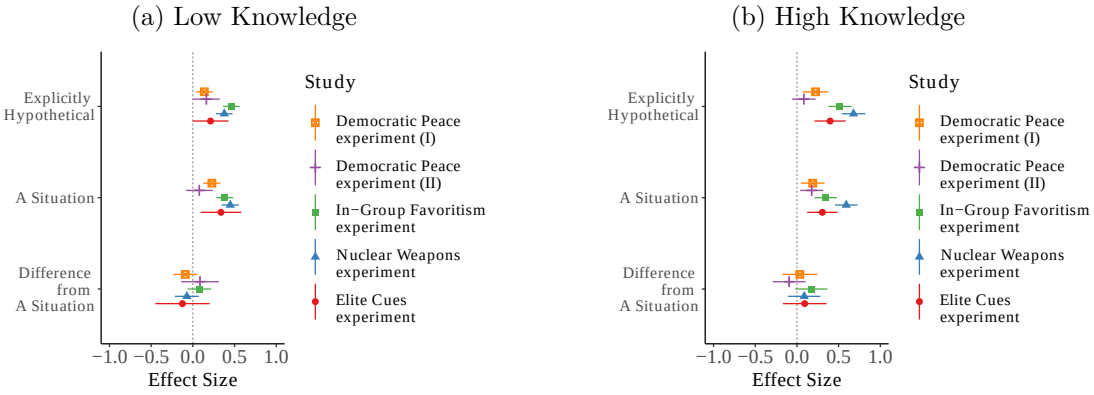

Figure A.18 shows the moderating effects of explicit (vs. implicit) hypotheticality, across different levels of survey respondents' political knowledge. All outcomes are standardized.

moderates main effects for respondents with either low or high political knowledge.

A.6 Strength of partisanship additional results

In this section, we report additional models in which we examine the extent to which strongly and weakly identifying partisans respond differently to situational hypothet-

Figure A.19: Moderating effects of calling events real by political knowledge

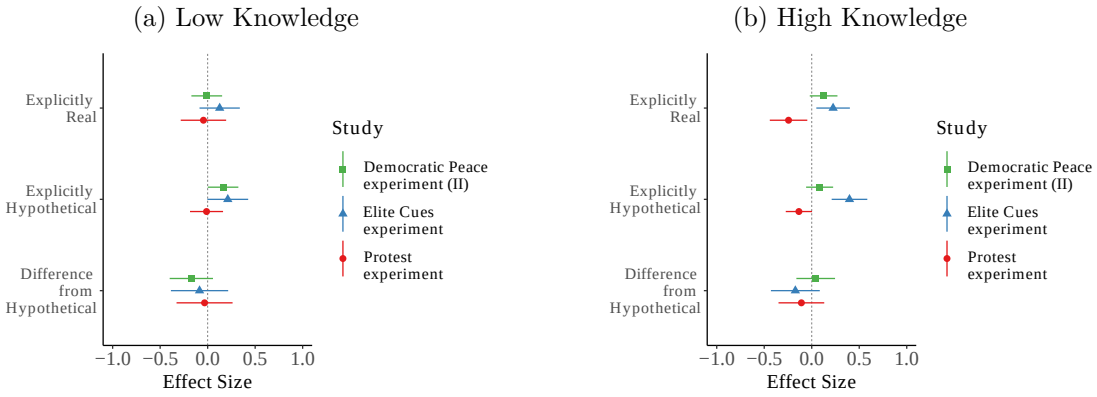

Figure A.19 shows the moderating effects of framing a scenario as real rather than hypothetical, across different levels of survey respondents' political knowledge. All outcomes are standardized.

icality. Specifically, in Figure A.20, we show that for the most part framing scenarios as explicitly real, or explicitly hypothetical does not moderate average treatment effects amongst strongly or weekly identifying partisans. One possible exception to this pattern, is strongly identifying partisans in the Elite Cues experiment. Indeed, it appears that framing a scenario as real in that experiment has a small negative effect which approaches conventional levels of statistical significance, slightly reducing the main effect of out-party cues on support for a given policy.

A.7 Study engagement: mobile versus computer

In this section, we consider how mobile and non-mobile users respond to different design choices relating to abstraction and detail. In Figure A.21, we report the moderating effect of our hypotheticality treatment, for mobile and non-mobile users. Our results suggest that whether a scenario is described as real or explicitly hypothetical does not moderate average treatment effects in our mobile and non-mobile subsamples.

In Figure A.22, we further consider the moderating effects of country names, for mobile and non mobile users, in the Democratic Peace experiment. We find little

Figure A.20: Moderating effects of calling events real by strength of partisanship

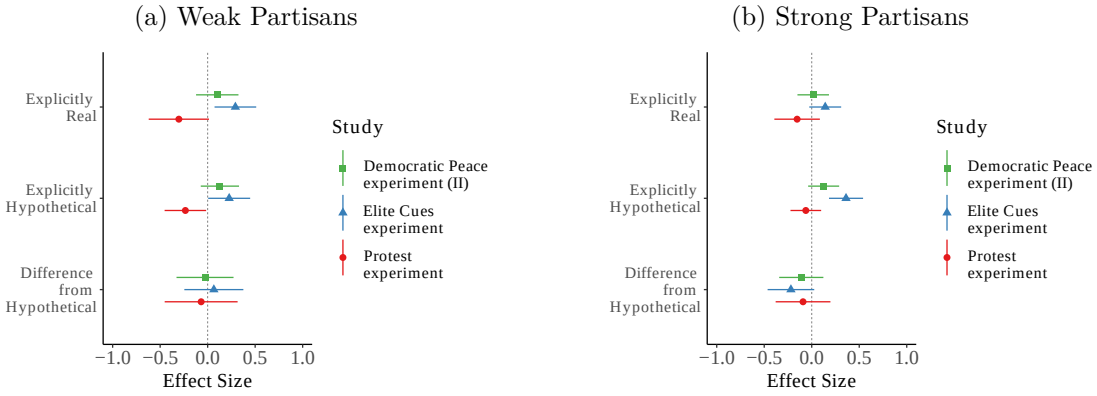

Figure A.20 shows the moderating effects of framing a scenario as real rather than hypothetical, across different levels of survey respondents' political knowledge. All outcomes are standardized.

Figure A.21: Moderating effects of situational hypotheticality by mobile use

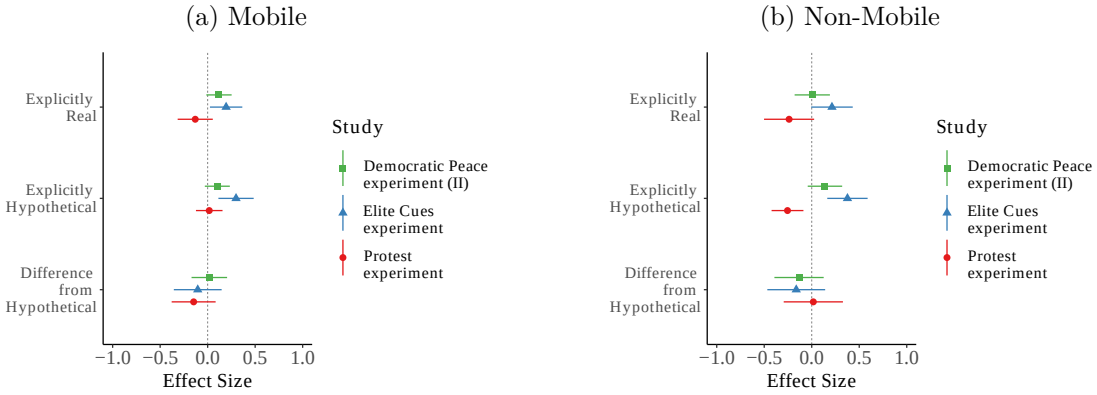

Figure A.21 shows the moderating effects of framing a scenario as explicitly real, rather than explicitly hypothetical, across subsamples of survey respondents using mobile and non-mobile devices. All outcomes are standardized.

Figure A.22: Moderating effects of country name by mobile use

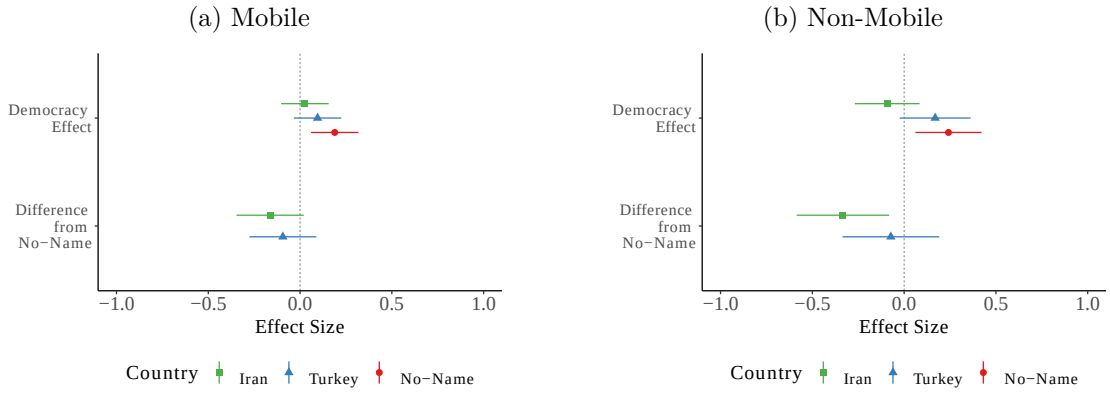

Figure A.22 shows the moderating effects of country name, across subsamples of survey respondents' using mobile and non-mobile devices, in the Democratic Peace experiment. All outcomes are standardized.

Figure A.23: Moderating effects of actor identity by mobile use

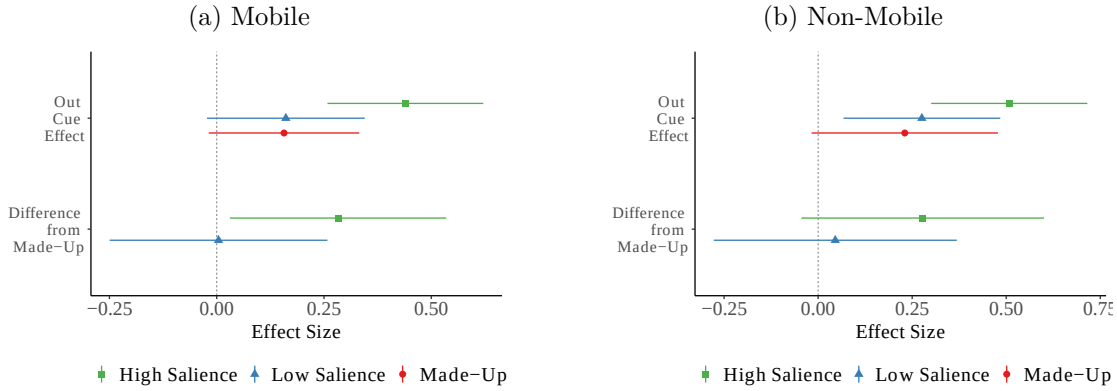

Figure A.23 shows the moderating effects of actor identity, across subsamples of survey respondents' using mobile and non-mobile devices, in the Elite Cues experiment. All outcomes are standardized.

evidence for that the consequences of country selection differ between our mobile and non-mobile subsample. Finally, in Figure A.23, we further demonstrate that the moderating effect of actor identity in the Elite Cues experiment is similar across our mobile and non-mobile samples.

These findings therefore suggest little evidence that questions of actor identity and situational hypotheticality have implications for mobile device usage, though as we

note in Chapter 7, the dimension of abstraction where mobile use would be the most relevant would be contextual detail, which we are unable to test for. We encourage future researchers to explore this question.

A.8 Do Different Dimensions of Abstraction and Detail Interact?

Throughout the book, we consider the moderating effects of design choices individually. However, one may wonder whether the consequences of different decisions regarding varying levels of design choices have interactive moderating effects on main treatments. To address this question, we leverage our Nuclear Weapons replication, in which we randomized both actor identity and contextual detail.

In Figure A.24, we present models where we consider the moderating effects of country names on original average treatment effects for two experimentally assigned sub-groups receiving either low or highly detailed vignettes. Generally, our findings suggest that the moderating effects of country names on original average treatment effects are not conditioned by the level of detail in an experimental vignette. However, we do find some evidence that adapting real world countries might have a small attenuating effect when context is low. That said, this conditional moderating effect, which approaches conventional levels of statistical significance ($p < 0.08$) will not lead experimenters to draw substantively different inferences.

To further investigate the cumulative effects of abstraction and detail along different dimensions of our framework, we created additive abstraction scores detailing the levels of abstraction and detail to which a subject was assigned (in any given vignette). This score is comprised of up to three dimensions: situational hypothetically, actor identity and contextual detail, depending on the type of abstraction manipulated in any given study. Higher values denote more detailed and realistic experiments.

For example, if a respondent was assigned to a Nuclear Weapons vignette which was described as explicitly hypothetical, and the vignette included an un-named

country and minimal context, than the respondent's corresponding abstraction score would be 0. Moving up in the ladder of detail on any one of our conceptual dimensions would increase this score. Thus, being assigned to a non-explicitly hypothetical vignette would increase the score by one point. Similarly, variation in our actor identity condition could increase the score by up to three points (because respondents were assigned to four conditions), and additional context can also increase the score by one point.

In Figure A.25, we test whether our abstraction scale moderates original ATEs. We find that overall levels of abstraction have a sharp null effect in our Elite Cues and Nuclear Weapons experiments. In addition, the scale has a modest albeit statistically significant attenuating effect on the ATE of our In-Group Favoritism and Democratic Peace experiments. Given the results reported in the main text, we expect this attenuation in the In-Group Favoritism experiment to be driven, largely, by additional context which reduces the dosage of original treatments vis-a-vis background information. Similarly, we expect the attenuation in the Democratic Peace experiment to be driven by treatment inconsistency, created by the description of Iran as a Democracy.

Figure A.24: Moderating Effects of Country Name by Contextual Detail Subsamples

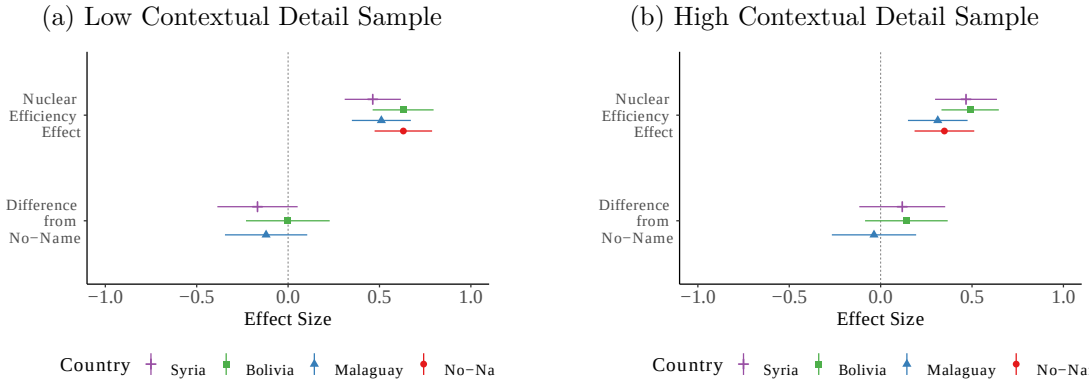

Figure A.24 shows that different country names do not moderate average treatment effects in diverging and substantively significant ways across low and high contextually detailed vignettes in the Nuclear Weapons experiment. In each panel, point estimates and corresponding confidence intervals are extracted from three separate OLS models where original outcomes are predicted by original treatments interacted with country names. In all models across both panels un-named countries are the reference category.

Figure A.25: Moderating Effects of Abstraction Scale on all ATEs

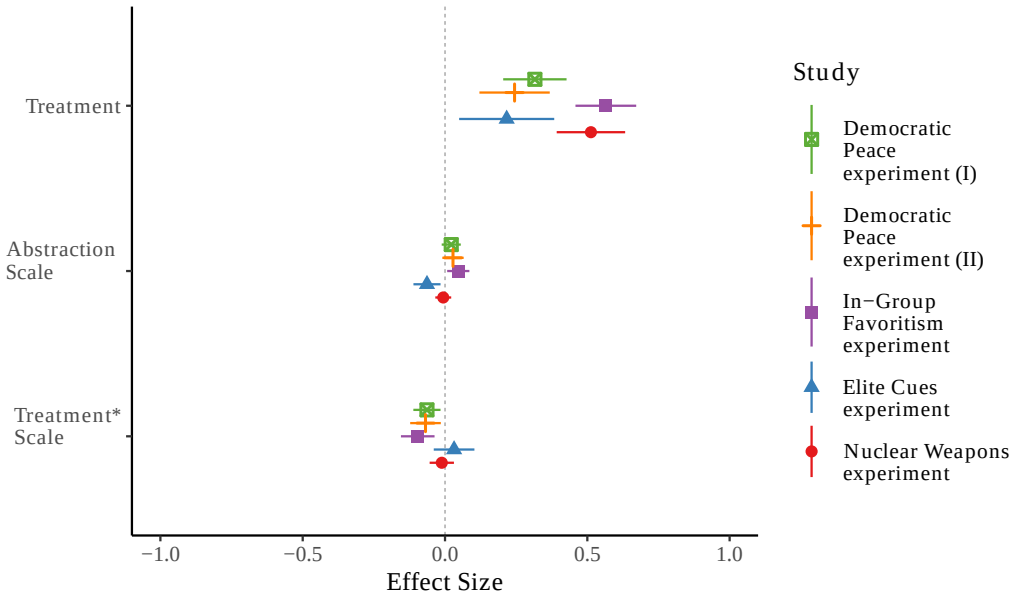

Figure A.25 demonstrates the limited moderating effects of our abstraction scale, on original ATEs. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with our abstraction scale.

Bibliography

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2021. "Beyond the breaking point? Survey satisficing in conjoint experiments." *Political Science Research and Methods* 9(1):53–71.

Brutger, Ryan and Brian Rathbun. 2021. "Fair Share?: Equality and Equity in American Attitudes towards Trade." *International Organization* 75(3):880–900.

Clifford, Scott and Jennifer Jerit. 2016. "Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions." *Public Opinion Quarterly* 80(4):858–887.

Coppock, Alexander and Oliver A McClellan. 2019. "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents." *Research & Politics* 6(1):2053168018822174.

Hill, Seth J and Gregory A Huber. 2019. "On the Meaning of Survey Reports of Roll-Call "Votes"." *American Journal of Political Science* 63(3):611–625.

Kam, Cindy D. 2012. "Risk Attitudes and Political Participation." *American Journal of Political Science* 56(4):817–836.

Kertzer, Joshua D. and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60(1):234–249.

Malhotra, Neil, Yotam Margalit and Cecilia Mo. 2013. "Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact." *American Journal of Political Science* 57(2):391–410.

Mutz, Diana C and Eunji Kim. 2017. "The impact of in-group favoritism on trade preferences." *International Organization* 71(4):827–850.

Nicholson, Stephen P. 2012. "Polarizing cues." *American journal of political science* 56(1):52–66.

Orr, Lilla V and Gregory A Huber. 2020. "The policy basis of measured partisan animosity in the united states." *American Journal of Political Science* 64(3):569–586.

Peyton, Kyle, Gregory A Huber and Alexander Coppock. 2021. "The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic." *Journal of Experimental Political Science* Forthcoming.

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.

Thomson, Keela S and Daniel M Oppenheimer. 2016. "Investigating an alternate form of the cognitive reflection test." *Judgment and Decision Making* 11(1):99.

Tomz, Michael and Jessica LP Weeks. 2020. "Public opinion and foreign electoral intervention." *American Political Science Review* 114(3):856–873.

Tomz, Michael R and Jessica LP Weeks. 2013. "Public opinion and the democratic peace." *American political science review* 107(4):849–865.