Abstraction and Detail in Experimental Design:
Supplementary appendix

Contents

A Survey Overview

The three experiments analyzed in our main text were embedded in two separate waves, implemented in Spring 2019, and Spring 2020. Specifically, our NUCLEAR WEAPONS and IN-GROUP FAVORITISM experiments were fielded in Spring 2019, followed by a second wave in Spring 2020 in which we fielded the ELITE CUE experiment. The implementation process of these studies followed a simple and common procedure further detailed in Figure A.1.

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 or are 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 whether no description of hypotheticality was mentioned in the introduction the experiment. This treatment varies 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 are 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 are randomly assigned versions of the original studies that vary in their amount of contextual detail, and 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, which we will employ in future analyses.

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

8. **Additional Demographic and individual difference batteries:** 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.

Figure A.1: Overview of Study Protocol

Our first survey, in which we embedded the NUCLEAR WEAPONS and IN-GROUP FAVORITISM experiments, were implemented with Dynata (formerly known as Survey Sampling International (SSI)).[1] In Table 1, we report descriptive statistics of our sample, including basic demographics, and all variables employed in our analyses. Our ELITE CUE study was embedded in a second survey, implemented with Lucid.[2] We present additional descriptive statistics for our Lucid sample in Table 2. Importantly, while the sampling strategies of both survey providers differ from one another, they both involve online panels. Future research should extend our findings on samples fielded using other sampling strategies, including on survey respondents that vary in their level of naïveté (e.g. Chandler, Mueller and Paolacci, 2014), as well as in other survey modes. One potential advantage of online survey experiments is that they afford researchers the opportunity to provide greater contextual detail than telephone-based survey experiments do, such that it is likely that contextual detail has greater attenuating effects in telephone-based survey experiments than it does online, for example.

Table 1: Descriptive Statistics - Study I (MK+PSV)

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

---

[1]For other recent studies in political science employing this platform for experimental research, see e.g. Kam (2012); Malhotra, Margalit and Mo (2013); Kertzer and Brutger (2016); Brutger and Rathbun (2020).

[2]Recent investigations suggest that Lucid is a suitable platform for implementing survey experiments in the U.S. context (Coppock and McClellan, 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). For additional political science studies implemented with Lucid, see Tomz and Weeks (2020); Hill and Huber (2019); Orr and Huber (2020).

Table 2: Descriptive Statistics - Study II (Nicholson)

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

B Study Instrumentation

B.1 CASE SELECTION

As Table 3 shows, the conceptual typology we employ in the piece for discussing different dimensions of abstraction is applicable to a wide range of experiments, which we illustrate below by coding a number of prominent experimental pieces in Table 3.

Table 3: Abstraction in experimental political science

		Type of abstraction		
Type of experiment	Example	Situational Hypotheticality	Actor Identity	Contextual Detail
Audit experiment	Butler and Broockman (2011)	Deception	N/A	Med
Conjoint experiment	Hainmueller and Hopkins (2015)	Implicit	Unnamed	Med
Econ-style lab experiment	Kanthak and Woon (2015)	Real	Unnamed	Low
Endorsement experiment	Lyall, Blair and Imai (2013)	Real	Real	Med
Framing experiment	Nelson, Clawson and Oxley (1997)	Deception	Real	High
Lab-in-the-field experiment	Habyarimana et al. (2007)	Real	Unnamed	Low
Scenario-based survey experiment	Tomz (2007)	Implicit	Unnamed	Med
War game	McDermott et al. (2007)	Simulation	Artificial	Med
Field Experiment	Lyall, Zhou and Imai (2020)	Real	Real	High

We believe our typology can be applied to any information provision experiment, where respondents are presented with information by the experimenter to see how it affects their behavior. In this article, we chose to investigate questions of abstraction in survey experiments in particular for two reasons. The first is their popularity. Survey experiments now constitute the most widely-used experimental method appearing in many top journals in political science: for example, the American Journal of Political Science published 205 experiments from 2011-2020, 121 of

which were survey experiments, pointing to the relevance of our focus here. The second concerns questions of tractability. Although issues of abstraction and concreteness in experimental design apply to non-survey experiments as well (such as economics-style bargaining games (e.g. Tingley and Walter, 2011; Kanthak and Woon, 2015; Kertzer and Rathbun, 2015)), they also introduce additional considerations. Economics-style lab experiments, for example, embrace abstract designs in order to isolate the effect of incentives – a consideration we set aside in the manuscript itself.

The manuscript ultimately replicates and/or extends three different survey experiments, chosen through three selection criteria. First, we chose prominent studies that focused on core theoretical debates across a range of subfields of political science. The ELITE CUES experiment comes from American politics (Nicholson, 2012), though the construct being studied — the role of elite endorsements in support for policy preferences – features prominently in the study of political behavior regardless of subfield (e.g. Bullock, 2011; Druckman, Peterson and Slothuus, 2013; Guisinger and Saunders, 2017; Bisgaard and Slothuus, 2018; McDonald, Croco and Turitto, 2019). The IN-GROUP FAVORITISM experiment comes from international political economy (Mutz and Kim, 2017), but questions of in-group favoritism and intergroup relations also loom large in the study of domestic politics as well (e.g. Iyengar and Westwood, 2015; McClendon, 2018; Nugent, 2020). The NUCLEAR WEAPONS experiment comes from international security, but the underlying questions it tests about the strength of our commitment to moral principles is by no means exclusively the preserve of the security literature (e.g. Chu, 2019; Ryan, 2019; Jung, 2020).

Second, because our quantity of interest is the interaction between a dimension of abstraction and the study-level treatment, the selected studies needed to have relatively simple designs to afford us sufficient statistical power. Complex or high-dimensional factorial experiments, or conjoint experiments with multiple treatments of interest, are less useful for our purposes than experiments that had only one factor of interest, particularly if the factor had only two or three levels.

Third, the experiments selected needed to demonstrate a large and substantively meaningful effect. If we were replicating or extending studies whose original treatments barely moved their outcome variable, the absence of heterogeneous treatment effects by levels of abstraction would be less informative than when replicating studies whose treatment effects were substantively large.

Fourth, the experiments needed to be conducive to manipulating our three dimensions of abstraction: situational hypotheticality, actor identity, and contextual detail. We did so in each experiment in different ways. For example, in the NUCLEAR WEAPONS experiment, whose stimuli are

relatively lengthy, we manipulated contextual detail by *cutting* context, whereas in the IN-GROUP FAVORITISM experiment, whose stimuli are relatively short, we were able to manipulate contextual detail by *adding* context. In the ELITE CUES experiment, we manipulate actor identity by manipulating the individual politician involved; in the NUCLEAR WEAPONS experiment we manipulate actor identity by manipulating the country involved. This also speaks to the value of selecting experiments from across multiple subfields of the discipline, since abstraction and concrete detail may manifest themselves in very different ways depending on the research question.

Finally, we note that the fact that our experiments differ from one another in a wide range of ways — they cover substantively different questions, from different quadrants of political science, were fielded on different survey platforms, at different points in time — and yet still converge on a common set of findings should add credibility to our findings, even though, as is the case with all social science research, additional studies are inevitably required, which we explore in additional work (Brutger et al., 2022).

B.2 ELITE CUES EXPERIMENT

The ELITE CUES experiment replicates and extends Nicholson's (2012) study of elite cues about immigration reform in the United States, to explore the effects of actor identity in experimental design.[3] Nicholson's original study examined the effect of in/out party endorsements on 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.[4] In each condition respondents were told whether the endorser was a Republican or Democrat and for the fictional endorser — Stephen Smith — the partisan affiliation was randomized. Respondents then indicated their support for the immigration reform policy. Following the main outcome variable, respondents were asked to think about the situation again then asked to complete a thought listing exercise and a factual manipulation check (whether the policy was endorsed by a member of a par-

---

[3]While Nicholson's study includes several experiments, considering different policies and cue-givers, we focus on the immigration policy experiment endorsed by politicians (rather than parties).

[4]Additionally, we update the substantive context of the experiment to focus on protection for "Dreamers" in the U.S.

ticular party or not endorsed by anyone). These latter questions enable us to determine how actor identities affect respondents comprehension and recall of the general experimental scenario as well as the treatment.

To replicate the main results presented in Nicholson (2012), all subjects read the introduction and vignette presented in Figure B.2, whose features randomly varied across respondents:[5]

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)

Figure B.2: Elite Cue Vignette

After administering our main outcome variable (shown at the bottom of Figure B.2) we asked respondents to complete a common thought listing task (See Figure B.3).

When you think about the situation / real situation / hypothetical situation you just read, what features of the situation / real situation / hypothetical situation 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 B.3: Thought Listing Exercise

Following the thought listing exercise in Figure B.3, we directly investigate respondents' attention to their main treatment condition, by employing a factual manipulation check (Kane and Barabas, 2019). To do so, we ask the question presented in Figure B.4:

---

[5]Note that underlined aquamarine text signifies our hypotheticality treatment, and *italicized blue text* signifies the original study's treatment, which we extended to include diverging types of actor identities (made up, low salience, high salience).

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 B.4: Elite Cue Treatment Manipulation Check

B.3 IN-GROUP FAVORITISM EXPERIMENT

The IN-GROUP FAVORITISM experiment replicates and extends portions of Mutz and Kim's (2017) investigation of American trade preferences, to study the effects of additional contextual detail. In replicating their basic framework, we focus on a common decision experimentalist 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.[6] 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.

In particular, when implementing our study, we consider how providing respondents with increased context moderates the main identified treatment effect. Thus 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. Apart from our contextual detail treatment, we follow a simplified version of the procedure implemented in Mutz and Kim (2017). In a similar fashion to our ELITE CUES study, we provide respondents with a thought listing exercise as well as a factual manipulation check. Doing so enables us to test whether increased contextual detail affects respondents' comprehension of experimental scenarios and treatments.[7]

To replicate and extend the main results of Mutz and Kim (2017), we present all subjects with the following introduction, along with a vignette whose contents randomly varied across respondents:

---

[6]Crucially, the distinction between filler and charged context is less about whether the additional context is relevant to the scenario or not – in most circumstances, experimentalists are unlikely to be interested in adding totally irrelevant text, which would present somewhat jarringly to respondents — and more about whether to add additional contextual information that they expect to interact with their treatment of interest.

[7]Future work should consider exploring whether charged context with numeric context produces different results than charged context without out numeric context, particularly for individuals at varying levels of numeracy (Mérola and Hitt, 2016).

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 / 10** people in the U.S. who gain a job and can now provide for their family, **10 / 1000** people in a country that we trade with will **gain new jobs and now be able to provide for their family / lose jobs and will no longer be able to provide for their family**.[a]

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: 1,000 - 10 - gain, 10 - 1,000 - gain, 10 - 1000 - lose.

Figure B.5: In-Group Favoritism Vignette

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

Based on the questions reported in Figure B.6, we created our main DV, measuring support for

- 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 B.6: In-Group Favoritism Outcomes

the described policy, on a four point scale ranging from strongly oppose to strongly support.

After collecting our main outcome variable we further ask respondents to engage in a thought listing task. The thought listing task is similar to the one reported in Figure B.3. Following the thought listing exercise detailed above, we directly investigate respondents' attention to their main treatment condition. To do so, we ask the following manipulation check reported in Figure B.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 B.7: In-Group Favoritism Manipulation Check

The NUCLEAR WEAPONS experiment replicates and 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.[8]

We used our replication to consider the joint effects of contextual detail and actor identity, adding two additional treatment arms to the original study on nuclear aversion. More specifically, 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 alternatives to country names, which include: (1) Syria (as in the original study), (2) an unnamed country ("a foreign country"), (3) a fictitious country name ("Malaguay"), or (4) a real and schema-inconsistent country (Bolivia). The extent to which real countries are schema-consistent with a given experimental scenario is an empirical question. Therefore, we fielded a pilot study on a sample of about 600 American adults recruited on Amazon Mechanical Turk, in which we described the experimental scenario in the NUCLEAR WEAPONS experiment in its un-named country format. We then presented the study's main outcome questions, and asked respondents to rate the likelihood that each of eleven different countries would be the actor in each scenario.[9] After the main outcome measure, we present respondents with a thought listing exercise and factual questions relating to the main treatment, as detailed in Appendix §2.3.

To replicate the main results in Press, Sagan and Valentino (2013), we present all subjects with the following text, as well as a summary table (see Table 4):

> 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.

---

[8]We simplified the original design to only include two treatment-levels for the probability of success, as as detailed in Appendix §2.3.

[9]For more information regarding our pretest procedure see Appendix §3.

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*

Table 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 question (see Figure B.8). Our main outcome of interest is approval of a nuclear weapon attack (the first item in Figure B.8).

- 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 B.8: Nuclear Weapons Outcome Questions

We further included a question from the original instrument, which is directed towards respondents who stated their preference for conventional attacks, and the reasons behind this selection. However, we do not analyze responses to this question in our paper. We also included a thought listing exercise relating to the nuclear weapons vignette, like the one depicted in Figure B.3. Lastly, we asked respondents a manipulation check question (see Figure B.9).

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 B.9: Nuclear Weapons Manipulation Check

15

C Power Calculations

In our experiments we have two sets of quantities of interest: the study-level treatment effects (e.g. in the NUCLEAR WEAPONS experiment, whether nuclear weapons are equally effective or dramatically more effective than conventional strikes), and interaction effects between the study-level treatments and our design treatments (e.g. whether the scenario is described as explicitly hypothetical or not). In order to ensure that these interaction effects are sufficiently powered, in this section, we consider the statistical power of our experimental design to detect theoretically meaningful moderating effects of different design choices. To do so, we focus on the NUCLEAR WEAPONS experiment, because it has the largest number of experimental cells, due to the fact that the country-name treatment includes four design-choice conditions: i) no-name, ii) made-up name, iii) schema-inconsistent name, and iv) schema-consistent name. In each of our main models, we compare the original study's average-treatment effect under the no-name condition, with one of the other country conditions. This effectively leads us to estimate models with approximately 1000 observations, in which our quantity of interest is the effect of nuclear effectiveness, conditional on country name choice.

Our key question is whether we are sufficiently powered to precisely estimate η, in the model depicted in equation 1. Specifically, we want to ensure that if altering country names in a given experiment (i.e. shifting γ_{design} from 0 to 1) shapes a study's average treatment effect, we would be sufficiently powered to detect it (formally denoted as $\eta(\beta_{treatment} * \gamma_{design})$).[10]

$$y_i = \alpha + \beta_{treatment} + \gamma_{design} + \eta(\beta_{treatment} * \gamma_{design}) + \epsilon_i \tag{1}$$

We use our data, as well as simulation procedure in the R package DeclareDesign to address this concern (Blair et al., 2019). Specifically, we declare a model by specifying three quantities: i) the average treatment effect of the nuclear weapon study (nuclear effectiveness), ii) the average treatment effect of a country name choice (describing the country as Syria rather than an unnamed country), and iii) the interaction between each treatment.

In Figure C.10, we report the main results from the DeclareDesign diagnosis based on 1,000

---

[10]Lenth (2007) provides a useful discussion of power analysis, which discourages the use of retrospective analyses. However, Lenth (2007, E26) also notes that retrospective power analysis can be valuable when it uses "what we have learned (e.g., the error SD) and an effect size deemed of clinical importance" to determine the appropriate amount of data to identify effects of importance. We do so in our power analysis, which shows that we are well powered to identify interaction effects that would offset a significant proportion of the main treatment effects.

simulations. We consider five different interaction estimands:

- A negative interaction effect of -0.04 – based on the coefficient from our original model, estimating the moderating effect of Syria country name.

- A negative interaction effect of -0.13 – Resembling an attenuation equivalent to a 25% decrease in the original study's ATE.

- A negative interaction effect of -0.27 – Resembling an attenuation equivalent to a 50% decrease in the original study's ATE.

- A negative interaction effect of -0.40 – Resembling an attenuation equivalent to a 75% decrease in the original study's ATE.

- A negative interaction effect of -0.54 – Resembling a full attenuation of the original study's ATE.
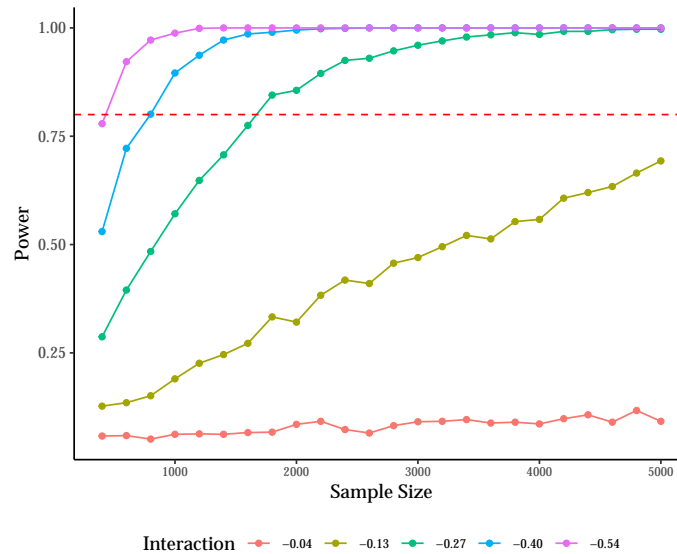
Figure C.10: Power Calculations

Figure C.10 demonstrates our power to detect different interaction effect sizes, conditional on sample size.

The results reported in Figure C.10 suggest that even with a sample of 400 subjects, we would be able to identify an interaction effect that fully attenuates our main treatment (Pink line). Note

17

that throughout the paper, all models include at least 1,000 subjects per comparison. Accordingly, we are relatively well-powered to detect moderating effects which attenuate our main average treatment effect by 50%-75% (blue and light-green line). That said, our ability to detect smaller attenuating effects — such as a 7%-25% attenuation in our main treatment is relatively limited (see dark green and orange lines).[11]

Overall, the results of this exercise are encouraging. Even in our models with the smallest number of observations, we are well powered to detect design-moderating-effects which would lead scholars to draw directionally different conclusions. More so, we are well powered to detect attenuating effects that reduce (increase) main effects substantively (i.e. halving or doubling the original effect size), without changing the direction of a given average treatment effect.

D   Pretest Procedure

On March 18, 2019 we fielded a survey on a sample of 600 American adults recruited using Amazon Mechanical Turk to test the schema consistency of 11 different countries with the experimental scenarios presented in the original Press, Sagan and Valentino (2013) study on US policy towards the development of nuclear attacks in foreign countries.[12]

Our survey started off by requesting informed consent and screening out respondents located outside the US or respondents accessing the survey through non-desktop devices. To ensure the comparability of our pre-test and main study, we randomized all original study-level treatments apart from country name which was held constant at the unnamed country condition. After completing the scenario respondents were presented with a matrix of eleven countries, and asked: "On a scale of 1-5, where 1 is very unlikely and 5 is very likely, How likely is it that the above scenario describes the following countries?" The countries included in our pre-test were:

Egypt, Iran, Ecuador, Bolivia, Sudan, Vietnam, Turkey, Ethiopia, Kyrgyzstan, Malaysia, Syria

Parallel analysis suggests the likelihood ratings load onto three factors; principal axis factoring with oblimin rotation suggests the following three clusters:[13]

---

[11]We also acknowledge that surveys with less attentive respondents would reduce the value of the results and undermine the power to analyze interaction effects.

[12]For recent articles fielded in political science journals using Amazon Mechanical Turk, see Brutger and Kertzer (2018); Tingley and Tomz (2014); Huff and Kertzer (2018); Renshon, Dafoe and Huth (2018).

[13]The model fit of a three-factor solution is good: RMSEA=0.055, TLI=0.963.

- **Countries outside the Middle East:** Ecuador, Bolivia, Vietnam, Ethiopia, Kyrgyzstan, Malaysia

- **Middle Eastern Adversaries:** Iran and Syria

- **Middle Eastern Others:** Egypt and Turkey

We therefore build on this clustering to inform our selection of country names, selecting Iran and Syria as schema consistent countries, and Ecuador and Bolivia as schema inconsistent countries.

*E   Actor Identities and Cognitive Burden and Treatment Recall*

In this section we present results of additional analyses relating to the ELITE CUE and NUCLEAR WEAPONS experiments. Specifically, in Figure E.11, we consider how the salience (and type) of an elite cue-giver, influences cognitive burden during the experimental procedure (measured by response latency). In Figure E.12, we further consider the effects of actor identity on treatment recall. Generally, we do not find evidence that actor type impacts cognitive burden or treatment recall in both the ELITE CUE and NUCLEAR WEAPONS experiments.

Figure E.11: Actor Identity Effects on Response Time
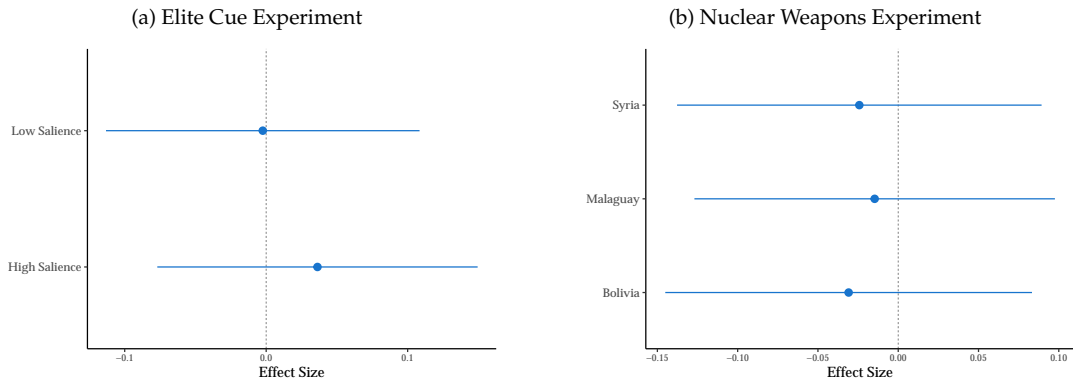
Panel (a) of Figure E.11 demonstrates that moving from a hypothetical actor to a low or high salience actor does not impact the cognitive burden of respondents (measured by logged response latency). Similarly, Panel (b) of Figure E.11 demonstrates that switching from an unnamed to a made up or real world country does not impact the cognitive burden of respondents (measured by response latency). Point estimates and corresponding confidence intervals are extracted from separate OLS models where the dependent variable (correctly responding to the treatment recall question), is regressed over the actor identity treatment.

Figure E.12: Actor Identity Effects on Treatment Recall

(a) Elite Cue Experiment

(b) Nuclear Weapons Experiment
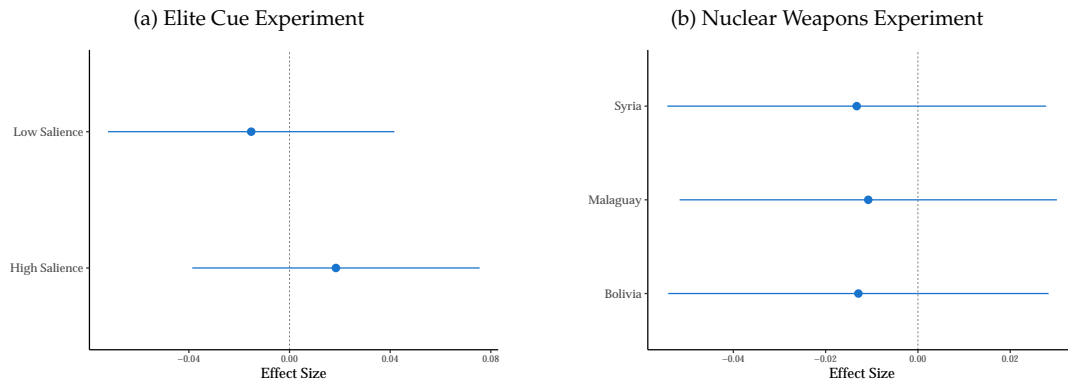
Panel (a) of Figure E.12 demonstrates that moving from a hypothetical actor to a low or high salience actor does not impact respondents' ability to correctly recall treatment. Similarly, panel (b) of Figure E.12 demonstrates that switching from an unnamed to a made up or real world country does not impact responses ability to correctly recall treatment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where the dependent variable (correctly responding to the treatment recall question), is regressed over the actor identity treatment.

In this Section, we consider how adding more context into vignettes affects response time. We find, in Figure F.13, strong evidence that longer vignettes increase cognitive burden measured by response latency. Increased cognitive burden, can further explain why respondents assigned to longer vignettes are less likely to correctly recall treatment (Figure 5 in the main text).

Figure F.13: Additional Context Effects on Response Time (NUCLEAR WEAPONS and IN-GROUP FAVORITISM Experiments)
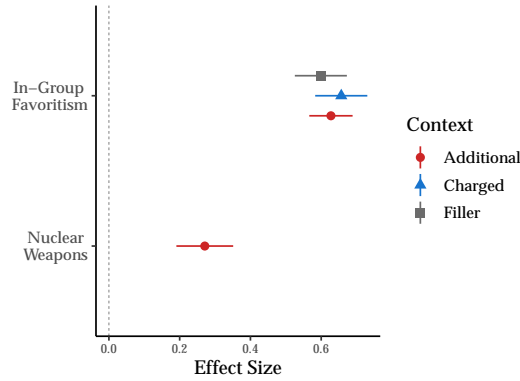

Figure F.13 demonstrates increasing context in experimental vignettes increases cognitive burden of respondents (measured by logged response latency). Point estimates and corresponding confidence intervals are extracted from separate OLS models where the dependent variable (response time for main outcome variable), is regressed over the actor identity treatment.

G   Situational Hypotheticality, Response Time, and Treatment Recall

In this section we examine whether situational hypotheticality affects response time and treatment recall success. To do so, we run additional models where we regress a measure of response time or treatment recall, over a binary variable taking the value of 1 if an experiment is introduced as explicitly hypothetical. Results reported in Figure G.14, suggest that situational hypotheticality does not affect response time in all three experiments. Results reported in Figure G.15 suggest that situational hyptheticality has a null effect on treatment recall in the ELITE CUES and NUCLEAR WEAPONS experiments. We do however, identify a small and marginally significant positive effect of situational hypotheticality on treatment recall in the IN-GROUP FAVORITISM experiment. Given the magnitude of this effect, and the fact that situational hypotheticality does not moderate average

treatment effects on our main outcomes, we suggest that varying levels of situational hypotheticality should not alter the substantive conclusions that researchers draw in their experimental studies.

Figure G.14: Situational Hypotheticality Effects on Response Time (ELITE CUES, IN-GROUP FAVORITISM, and NUCLEAR WEAPONS Experiments)
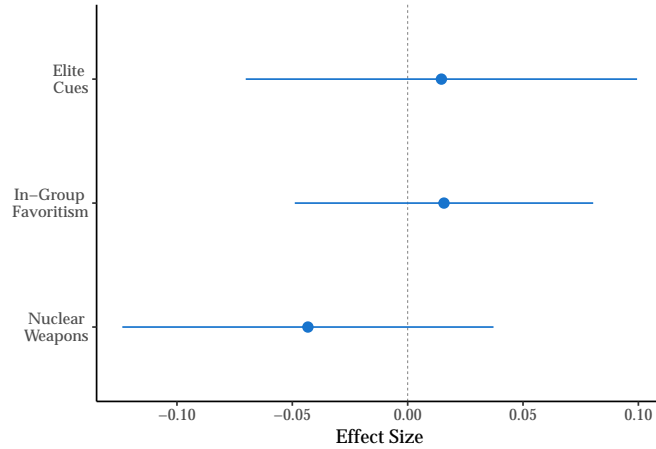

Figure G.14 demonstrates that introducing an experimental vignette as explicitly hypothetical does not affect the cognitive burden of respondents (measured by logged response latency). Point estimates and corresponding confidence intervals are extracted from separate OLS models where the dependent variable (response time for main outcome variable), is regressed over the situational hypotheticality treatment.

H Do Different Dimensions of Abstraction and Detail Interact?

Throughout the paper, 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 H.16, 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 moderat-

22

Figure G.15: Situational Hypotheticality Effects on Treatment Recall (ELITE CUES, IN-GROUP FAVORITISM, and NUCLEAR WEAPONS Experiments)
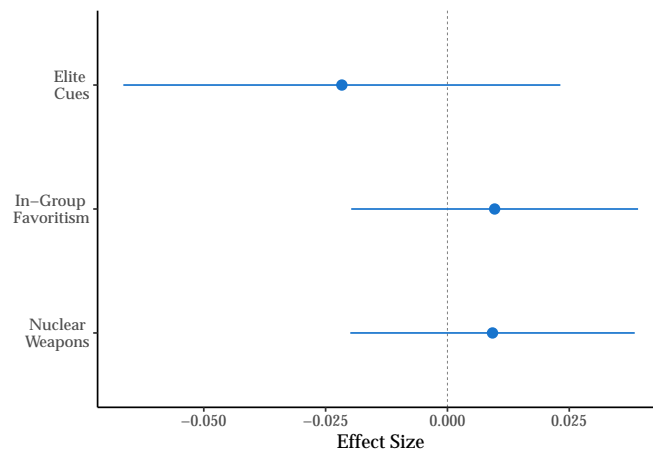

Figure G.15 demonstrates that introducing an experimental vignette as explicitly hypothetical does not affect treatment recall in the ELITE CUES and NUCLEAR WEAPONS experiments, but has a small and marginally significant positive effect on treatment recall in the IN-GROUP FAVORITISM experiment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where the dependent variable (correct treatment recall), is regressed over the situational hypotheticality treatment.

ing effect, which approaches conventional levels of statistical significance ($p < 0.08$) will not lead experimenters to draw substantively different inferences.

To further investigate the additive effect of abstraction and detail along different dimensions of our framework, we crated 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 respondents' 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 H.17, 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 experiment. 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.

Figure H.16: Moderating Effects of Country Name by Contextual Detail Subsamples

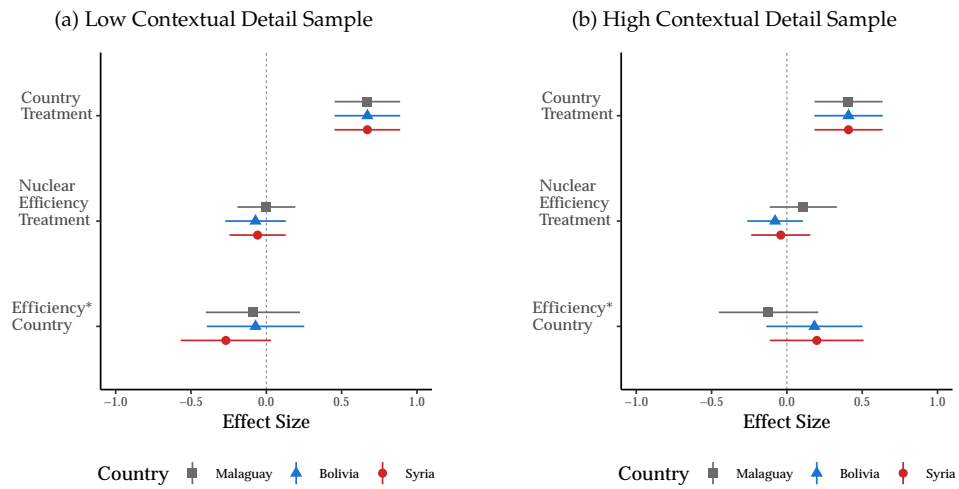(a) Low Contextual Detail Sample   (b) High Contextual Detail Sample

Figure H.16 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.

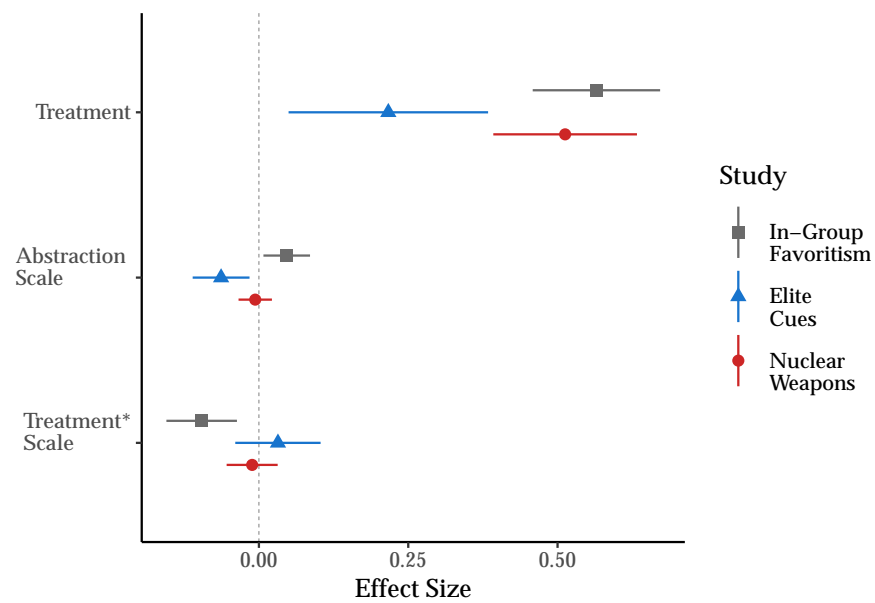Figure H.17: Moderating Effects of Abstraction Scale on all ATEs

Figure H.17 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.

In this section, we explore the extent to which different types of survey respondents react differently to levels of abstraction and detail in experimental design. To do so, we focus on two individual differences of theoretical relevance to the study of survey responses: political knowledge, and need for cognition. For ease of interpretation in the analysis below, we re-estimate our models from the main text on separate subsamples of respondents, mean-splitting by levels of political knowledge, and need for cognition, respectively. Although this facilitates ease of interpretation, it also reduces our sample size in each analysis, such that we encourage readers to take some caution when interpreting these additional results. In general, though, we find stronger results for political knowledge than we do for need for cognition. In particular, we find that contextual detail attenuates treatment effects for both high and low-knowledge respondents alike, and that high knowledge respondents are perhaps more sensitive to the use of high salience actors, but find few consistent patterns of differences between low and high-cognition respondents.

I.1 POLITICAL KNOWLEDGE

In our NUCLEAR WEAPONS and IN-GROUP FAVORITISM experiments, we measured political knowledge with two multiple choice questions regarding: i) the identity of the United Kingdom's current prime-minister, and ii) the length of U.S. House of Representative terms for office. In the ELITE CUES experiment, we added a third question regarding the identity of Israel's current prime-minister. Based on these questions, we split our sample in two based on whether respondents scored above the mean level of political knowledge.

In Figure I.18, we report results from models which consider the moderating effects of context in the IN-GROUP FAVORITISM experiment, amongst two 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. The charged context has a slightly stronger negative effect, and attains statistical significance among high knowledge respondents. In Figure I.19, we conduct the same exercise for the NUCLEAR WEAPONS experiment. Again, 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.

Figure I.18: Moderating Effects of Context By Political Knowledge in In-Group Favoritism Experiment
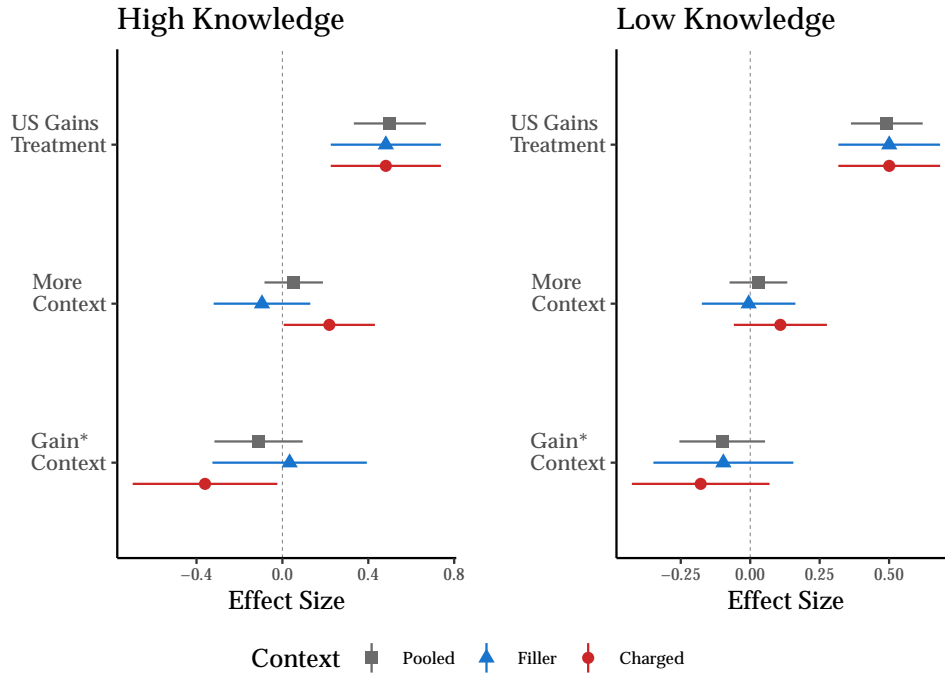

Figure I.18 demonstrates similar patterns for the moderating effect of context on original ATEs in the IN-GROUP FAVORITISM experiment between high and low-knowledge respondents. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

We therefore obtain relatively similar findings for the moderating effects of contextual detail between respondents high and low in political knowledge. In Figure I.20, we shift towards actor identity, examining the moderating effect of country name in the NUCLEAR WEAPONS experiment amongst our two subsamples. We find that for high political knowledge subjects, all country names (compared with an unnamed country) do not significantly moderate original average treatment effects. A somewhat similar pattern emerges when focusing on subjects with low political knowledge. Nonetheless, it appears that employing Bolivia as a country (rather than an unnamed country), has a small and unexpectedly positive moderating impact on the original average treatment effect for low knowledge respondents, though given the small sample size, some caution should once again be taken in interpreting the result.

28

Figure I.19: Moderating Effects of Context By Political Knowledge in Nuclear Weapons Experiment
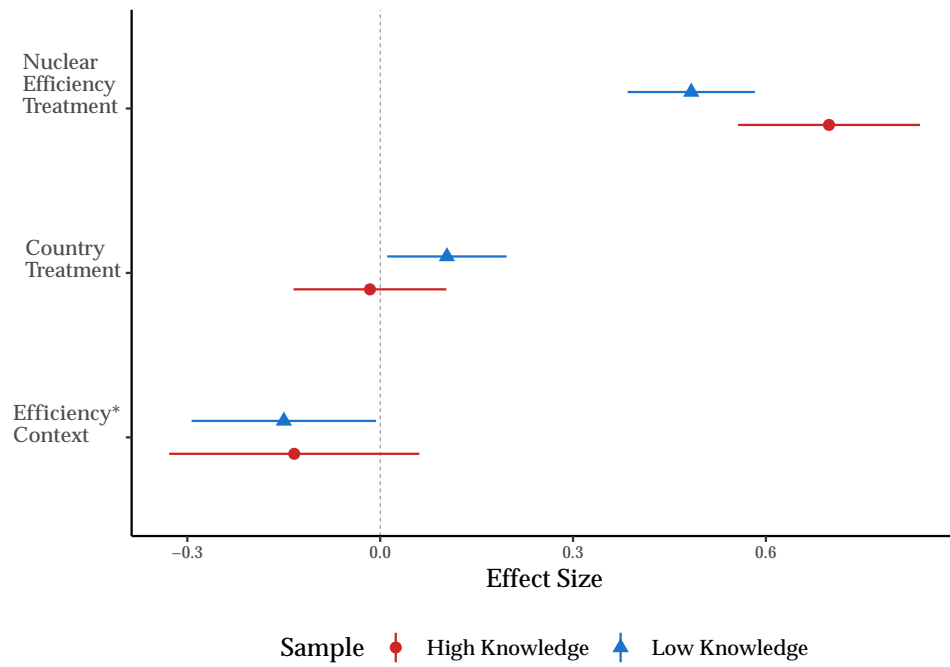

Figure I.19 demonstrates similar patterns for the moderating effect of context on original ATEs in the NUCLEAR WEAPONS experiment between high and low-knowledge respondents. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

Figure I.20: Moderating Effects of Country By Political Knowledge in Nuclear Weapons Experiment

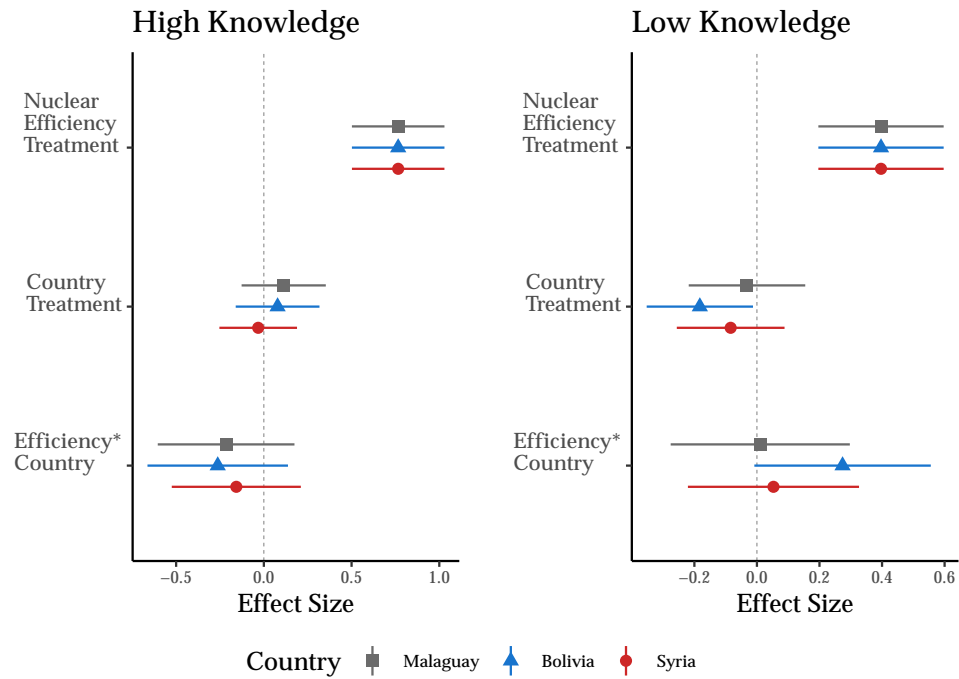Figure I.20 demonstrates slight differences of the moderating effect of some country names on original ATEs, when focusing on two samples of subjects with low and high political knowledge in the NUCLEAR WEAPONS experiment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

In Figure I.21, we continue examining the moderating effects of actor identity, this time focusing on the moderating effect of high and low salience actors, amongst subjects with low and high levels of political knowledge. As a reminder, in our main analysis, we find that employing high salience actors has a positive moderating effect, increasing the size of the average treatment effects. When splitting our samples, we do not find much evidence that employing low or high salience actors (compared with made up actors) moderates effects amongst low knowledge subjects. However, for higher knowledge respondents, who presumably have stronger priors about real-world political figures, the employing high salience actors does have a positive and statistically significant moderating effect.

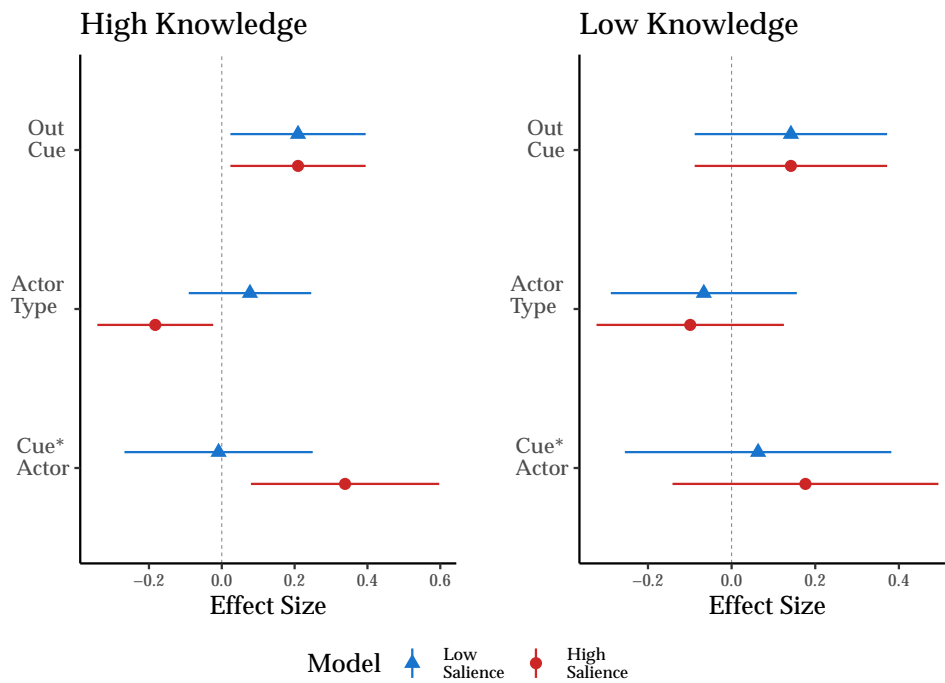Figure I.21: Moderating Effects of Actor Identity By Political Knowledge in Elite Cue Experiment

Figure I.21 suggests the persuasive effect of cues from high salient actors may be driven by higher knowledge respondents. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

I.2 NEED FOR COGNITION

We next examine our results amongst two subsets of respondents with high and low levels of need for cognition. To do so, we utilized a shorter-form version of the need for cognition scale based on 14 commonly used questions (Cacioppo and Petty, 1982; Rathbun, Kertzer and Paradis, 2017). We mean-split this index to create two subsamples of respondents, based on whether they display high (above average) or low (below average) levels of need for cognition. Because we did not the need for cognition item in the dispositional battery that accompanied the *Elite Cue* experiment, our analysis below focuses onthe IN-GROUP FAVORITISM and NUCLEAR WEAPONS Experiments.

In Figure I.22, we consider whether the moderating effects of context in the IN-GROUP FA-VORITISM experiment, vary between respondents with low and high levels of need for cognition. We find that varying levels of context does not affect respondents with high need for cognition. However, amongst respondents with low need for cognition, additional context, especially when charged, seems to attenuate average treatment effects. When turning to the NUCLEAR WEAPONS experiment in Figure I.23, we find that additional context negatively moderates average treatment effects, but this moderating effect is statistically significant only for respondents with higher levels of need for cognition. Given the mixed findings reported from the IN-GROUP FAVORITISM and NUCLEAR WEAPONS experiments, we suggest that readers take these results with a grain of salt, and encourage future research to further examine the relationship between need for cognition and context.

Finally, in Figure I.24 we consider the extent to which the moderating effect of country name varies across our subsamples of respondents with low and high levels of need for cognition. We once again find little evidence for a consistent pattern. Indeed, it appears that country names are unlikely to moderate average treatment effects for respondents with low and high need for cognition.

Figure I.22: Moderating Effects of Context By Need For Cognition in In-Group Favoritism Experiment

Figure I.22 demonstrates slight differences of the moderating effect of additional context on original ATEs, when focusing on two samples of subjects with low and high need for cognition in the IN-GROUP FAVORITISM experiment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

Figure I.23: Moderating Effects of Context By Need For Cognition in In-Group Nuclear Weapons Experiment

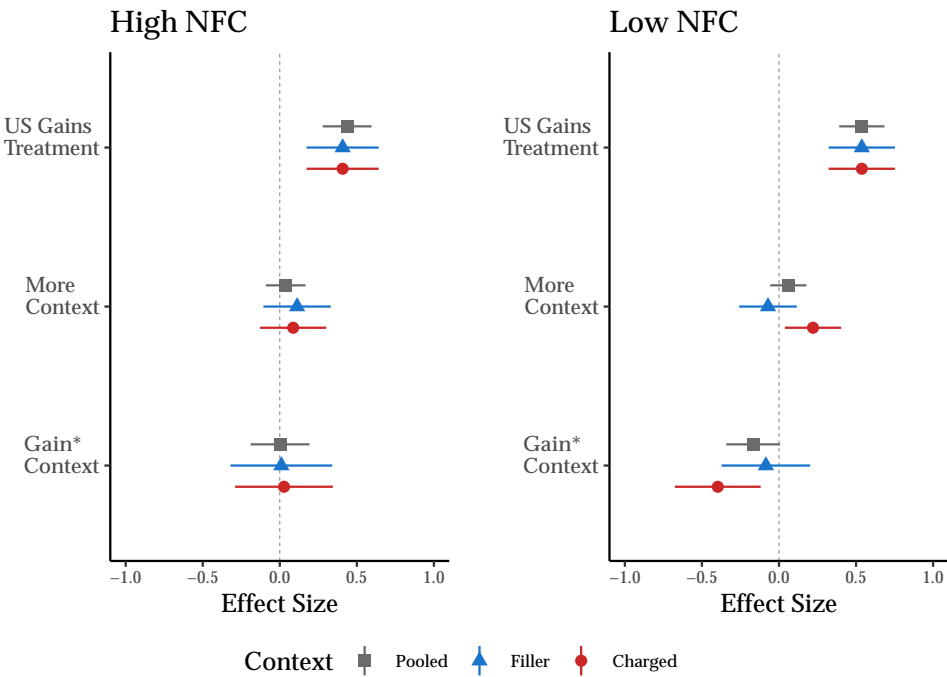Figure I.23 demonstrates slight differences of the moderating effect of additional context on original ATEs, when focusing on two samples of subjects with low and high need for cognition in the NUCLEAR WEAPONS experiment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator.

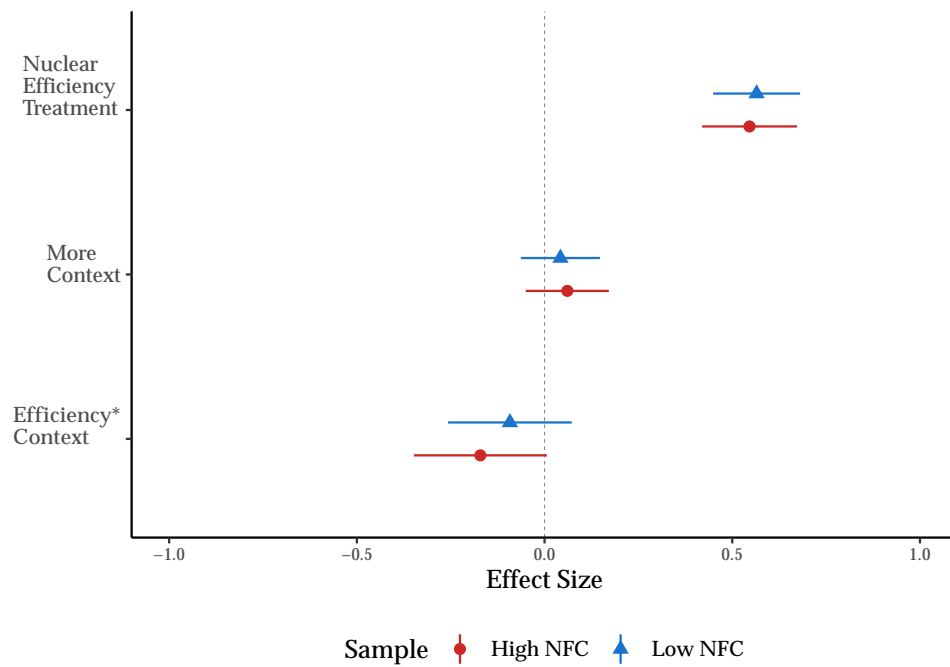Figure I.24: Moderating Effects of Country By Need For Cognition in In-Group Nuclear Weapons Experiment
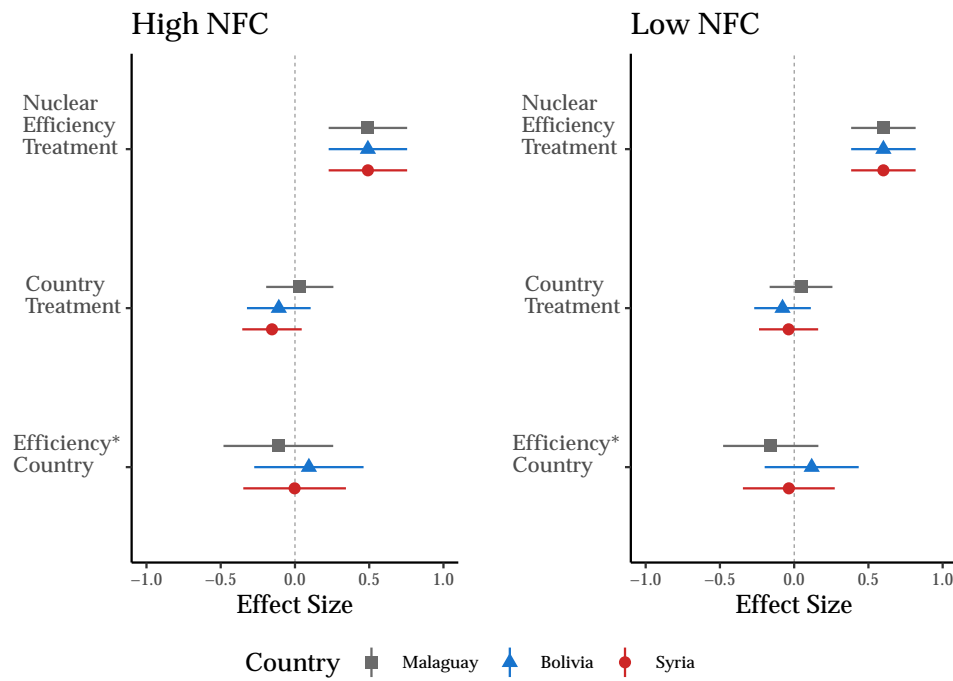
Figure I.24 demonstrates limited differences of the moderating effect of country on original ATEs, when focusing on two samples of subjects with low and high need for cognition in the NUCLEAR WEAPONS experiment. Point estimates and corresponding confidence intervals are extracted from separate OLS models where original outcomes are regressed over study treatments interacted with a context indicator. The high need for cognition sample includes respondents whose score on a need for cognition scale is above average, whereas low need for cognition sample includes respondents whose score on a need for cognition scale is below average.

Table 5: Replication of ATEs from Three Experiments

	Elite Cues		In-Group Favoritism	Nuclear Weapons	
	Model 1	Model 2	Model 3	Model 4	Model 5
Out-party Cue	0.25*				
	(0.06)				
Out-party Cue (Original)		0.32			
		(0.17)			
U.S. Gains			0.50*		
			(0.05)		
Nuclear Effectiveness				0.47*	
				(0.09)	
Nuclear Effectiveness (Original)					0.57*
					(0.11)
Num. obs.	1151	240	1507	535	319

$^{*}p < 0.05$. In Replication Study we analyze a subset of data that resembles the abstraction level of the original study.

J Regression tables

To preserve space, in the main text we present our results graphically; the companion regression tables are presented below. First, in Table 5 we present results from Figure 1, where we replicate the results from the NUCLEAR WEAPONS, ELITE CUES, and IN-GROUP FAVORITISM experiments. Second, in Table 6, we present results from Figure 2, considering the moderating effect of situational hypotheticality on original average treatment effects of all studies. Third, in Table 7, we present results from Figure 3 of the main text, considering the moderating effects of actor identity in the NUCLEAR WEAPONS and ELITE CUES experiment. Fourth, in Table 8, we report results from Figure 4 in the main text, which considers how additional context attenuates average treatment effects in the NUCLEAR WEAPONS and IN-GROUP FAVORITISM experiments. Finally, in Table 9, we present results from Figure 5 of the main text, which show how additional context in experimental vignettes reduces success in treatment recall questions.

Table 6: No moderating effects of situational hypotheticality

	Elite Cues	In-Group Favoritism	Nuclear Weapons
Out-party Cue	0.19*		
	(0.07)		
U.S. Gains		0.38*	
		(0.04)	
Nuclear Effectiveness			0.50*
			(0.04)
Hypothetical	0.11	−0.04	0.03
	(0.07)	(0.04)	(0.04)
Cue*Hypothetical	0.13		
	(0.10)		
Gain*Hypothetical		0.09	
		(0.06)	
Effectiveness*Hypothetical			−0.03
			(0.06)
Num. obs.	1633	4491	4462

$^{*}p < 0.05$. In the elite cue experiment, we omit respondents who were assigned to a baseline condition, where scenarios weren't described as hypothetical or real.

Table 7: Moderating effects of actor identity condition

	Elite Cues		Nuclear Weapons		
	Model 1	Model 2	Model 3	Model 4	Model 5
Out-group Cue	0.19*	0.19*			
	(0.07)	(0.07)			
Low Salience	0.02	0.02			
	(0.07)	(0.07)			
High Salience		−0.14*			
		(0.07)			
Cue*Low Salience	0.01	0.01			
	(0.10)	(0.10)			
Cue*High Salience		0.27*			
		(0.10)			
Nuclear Effectiveness			0.55*	0.55*	0.55*
			(0.08)	(0.08)	(0.08)
Malaguay			0.04		
			(0.07)		
Bolivia				−0.07	
				(0.07)	
Syria					−0.05
					(0.07)
Effective*Malaguay			−0.10		
			(0.12)		
Effective*Bolivia				0.05	
				(0.12)	
Effective*Syria					−0.04
					(0.11)
Num. obs.	1622	2398	1159	1122	1167

$^{*}p < 0.05$

Table 8: Adding contextual detail attenuates treatment effects

| | In-Group Favoritism | | | Nuclear Weapons |
	Model 1	Model 2	Model 3	Model 4
U.S. Gains	0.51*	0.50*	0.50*	
	(0.07)	(0.05)	(0.05)	
Filler	−0.01			
	(0.07)			
Charged		0.10*		
		(0.05)		
Pooled			0.05	
			(0.04)	
Gain*Filler	−0.07			
	(0.10)			
Gain*Charged		−0.20*		
		(0.07)		
Gain*pooled			−0.11	
			(0.06)	
Nuclear Effectiveness				0.56*
				(0.04)
Additional Context				0.06
				(0.04)
Effectiveness*Context				−0.15*
				(0.06)
Num. obs.	1511	2982	4491	4462

$^*p < 0.05$. In models 1-2, we compare charged and filler to a control condition. In model 3 we pool both conditions, and compare to control condition.

Table 9: Contextual Detail Effects on Treatment Recall Success

| | In-Group Favoritism | | Nuclear Weapons |
	Model 1	Model 2	Model 3
Charged	−0.13*		
	(0.02)		
Filler		−0.08*	
		(0.02)	
Additional Context			−0.03*
			(0.01)
Num. obs.	2946	2971	4395

$^*p < 0.05$

References

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.

Bisgaard, Martin and Rune Slothuus. 2018. "Partisan elites as culprits? How party cues shape partisan perceptual gaps." *American Journal of Political Science* 62(2):456–469.

Blair, Graeme, Jasper Cooper, Alexander Coppock and Macartan Humphreys. 2019. "Declaring and diagnosing research designs." *American Political Science Review* 113(3):838–859.

Brutger, Ryan and Brian Rathbun. 2020. "Fair Share?: Equality and Equity in American Attitudes towards Trade." *International Organization* Forthcoming.

Brutger, Ryan and Joshua D. Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72(3):693–724.

Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon and Chagai M. Weiss. 2022. "Abstraction in Survey Experiments: Testing the Tradeoffs." Book manuscript.

Bullock, John G. 2011. "Elite Influence on Public Opinion in an Informed Electorate." *American Political Science Review* 105(3):496–515.

Butler, Daniel M. and David E. Broockman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55(3):436–477.

Cacioppo, John T and Richard E Petty. 1982. "The need for cognition." *Journal of personality and social psychology* 42(1):116.

Chandler, Jesse, Pam Mueller and Gabriele Paolacci. 2014. "Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers." *Behavior research methods* 46(1):112–130.

Chu, Jonathan A. 2019. "A clash of norms? How reciprocity and international humanitarian law affect American opinion on the treatment of POWs." *Journal of Conflict Resolution* 63(5):1140–1164.

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.

Druckman, James N, Erik Peterson and Rune Slothuus. 2013. "How elite partisan polarization affects public opinion formation." *American Political Science Review* 107(1):57–79.

Guisinger, Alexandra and Elizabeth N. Saunders. 2017. "Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion Across International Issues." *International Studies Quarterly* 61(2):425–441.

Habyarimana, James, Macartan Humphreys, Daniel N Posner and Jeremy M Weinstein. 2007. "Why does ethnic diversity undermine public goods provision?" *American Political Science Review* 101(4):709–725.

Hainmueller, Jens and Daniel J Hopkins. 2015. "The hidden american immigration consensus: A conjoint analysis of attitudes toward immigrants." *American Journal of Political Science* 59(3):529–548.

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.

Huff, Connor and Joshua D. Kertzer. 2018. "How the Public Defines Terrorism." *American Journal of Political Science* 62(1):55–71.

Iyengar, Shanto and Sean J Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science* 59(3):690–707.

Jung, Jae-Hee. 2020. "The Mobilizing Effect of Parties' Moral Rhetoric." *American Journal of Political Science* 64(2):341–355.

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

Kane, John V and Jason Barabas. 2019. "No harm in checking: Using factual manipulation checks to assess attentiveness in experiments." *American Journal of Political Science* 63(1):234–249.

Kanthak, Kristin and Jonathan Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." *American Journal of Political Science* 59(3):595–612.

Kertzer, Joshua D. and Brian C. Rathbun. 2015. "Fair is Fair: Social Preferences and Reciprocity in International Politics." *World Politics* 67(4):613–655.

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.

Lenth, Russell V. 2007. "Statistical power calculations." *Journal of animal science* 85(suppl_13):E24–E29.

Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107(4):679–705.

Lyall, Jason, Yang-Yang Zhou and Kosuke Imai. 2020. "Can Economic Assistance Shape Combatant Support in Wartime? Experimental Evidence from Afghanistan." *American Political Science Review* 114(1):126–143.

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.

McClendon, Gwyneth H. 2018. *Envy in politics*. Princeton, NJ: Princeton University Press.

McDermott, Rose, Dominic Johnson, Jonathan Cowden and Stephen Rosen. 2007. "Testosterone and Aggression in a Simulated Crisis Game." *Annals of the American Academy of Political and Social Science* 614(1):15–33.

McDonald, Jared, Sarah E Croco and Candace Turitto. 2019. "Teflon Don or Politics as Usual? An Examination of Foreign Policy Flip-Flops in the Age of Trump." *The Journal of Politics* 81(2):757–766.

Mérola, Vittorio and Matthew P Hitt. 2016. "Numeracy and the persuasive effect of policy information and party cues." *Public Opinion Quarterly* 80(2):554–562.

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

Nelson, Thomas E., Rosalee A. Clawson and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91(3):567–583.

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

Nugent, Elizabeth. 2020. "The Psychology of Repression and Polarization." *World Politics* 72(2):291–334.

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* .

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.

Rathbun, Brian C, Joshua D Kertzer and Mark Paradis. 2017. "Homo diplomaticus: Mixed-method evidence of variation in strategic rationality." *International Organization* pp. S33–S60.

Renshon, Jonathan, Allan Dafoe and Paul Huth. 2018. "Leader Influence and Reputation Formation in World Politics." *American Journal of Political Science* 62(2):325–339.

Ryan, Timothy J. 2019. "Actions versus consequences in political arguments: Insights from moral psychology." *The Journal of Politics* 81(2):426–440.

Thomson, Keela S and Daniel M Oppenheimer. 2016. "Investigating an alternate form of the cog-

nitive reflection test." *Judgment and Decision Making* 11(1):99.

Tingley, Dustin H and Barbara F Walter. 2011. "The effect of repeated play on reputation building: an experimental approach." *International Organization* 65(2):343–365.

Tingley, Dustin and Michael Tomz. 2014. "Conditional Cooperation and Climate Change." *Comparative Political Studies* 47(3):344–368.

Tomz, Michael. 2007. "Domestic audience costs in international relations: An experimental approach." *International Organization* 61(4):821–840.

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