

Abstraction and Detail in Experimental Design

October 16, 2020

Ryan Brutger¹, Joshua D. Kertzer², Jonathan Renshon³, Dustin Tingley⁴ & Chagai M. Weiss⁵

ABSTRACT: Experimentalists in political science often face the question of how abstract or concrete their experimental stimuli should be. Typically, this question is framed in terms of tradeoffs relating to experimental control and generalizability: the more context you introduce into your studies, the less control you have, and the more difficulty you have generalizing your results. Yet we have reasons to question this framing of the tradeoff, and there is relatively little systematic evidence experimenters can rely on when calibrating the degree of abstraction in their studies. We seek to make two contributions with this project. First, we provide a theoretical framework which identifies and considers the consequences of three dimensions of abstraction in experimental design: situational hypotheticality, actor identity, and contextual detail. Second, we replicate a range of classic vignette-based survey experiments from political science, varying these levels of abstraction. Our results suggest, that apart from a specific set of conditions, there are fewer tradeoffs between abstraction and detail in survey experiment design than political scientists often assume.

ACKNOWLEDGMENTS: This paper benefited from helpful feedback from audiences at MIT's Political Experiments Research Lab in 2014, MPSA 2019, APSA 2020 and the 2020 NYU Experiments Conference, Adam Berinsky, Adam Seth Levine, Jonathan Mummolo, Rich Nielsen, Anne Sartori, Jonathan Woon, and Teppei Yamamoto. Authors listed in alphabetical order.

Word count: 8363 words in text + 2829 in bibliography + 686 in captions = 11841 words

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

²Paul Sack Associate Professor of Political Economy, Department of Government, Harvard University. Email: jkertzer@gov.harvard.edu. Web: <http://people.fas.harvard.edu/~jkertzer/>

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

⁴Professor, Department of Government, Harvard University. Email: dtingley@gov.harvard.edu. Web: <https://scholar.harvard.edu/dtingley>

⁵PhD Candidate, Department of Political Science, University of Wisconsin-Madison, Email cmweiss3@wisc.edu, Web: <http://chagaimweiss.com>

Experimentalists in political science often face a question about how abstract or concrete their experimental stimuli should be. This question is typically thought of in terms of tradeoffs relating to experimental control and generalizability. At one end of the continuum, researchers in the economics tradition for example, tend to prefer highly stylized experiments that are deliberately light on context, even though this comes at the expense of ecological validity and mundane realism (Morton and Williams, 2010, 313-14). This tradition is particularly popular in behavioral experiments in political science seeking to test the predictions of formal models (e.g., Dickson, 2009; Dawes, Loewen and Fowler, 2011; Tingley and Walter, 2011a,b; Kanthak and Woon, 2015; Kertzer and Rathbun, 2015; LeVeck and Narang, 2017; Quek, 2017; Hundley, 2020), but also arises in survey experimental work in political science as well (e.g. Renshon, 2015; Mutz and Kim, 2017; Tingley, 2017).

At the other end of the continuum, a different tradition originating in psychology has encouraged the use of rich and detailed vignette-based experiments. This approach, too, has found a foothold in political science (e.g., Rousseau and Garcia-Retamero, 2007; Brooks and Valentino, 2011; Druckman, Peterson and Slothuus, 2013; Teele, Kalla and Rosenbluth, 2018; Reeves and Rogowski, 2018; Bracic and Murdie, 2019; Tomz, Weeks and Yarhi-Milo, 2020). Rich and detailed stimuli are in some ways a response to the “major problem in public opinion and survey research”: the “ambiguity that often arises when survey respondents are asked to make decisions and judgments from rather abstract and limited information” (Alexander and Becker, 1978, 103). The ability to generalize experimental findings to other contexts, and the degree to which an experiment triggers the psychological process that would occur in the “real world”, are both thought to rise in proportion to the level of “realism” in a given vignette (Aguinis and Bradley, 2014, 361). Similarly, others argue that “concrete, realistic context” results in more “reliable assessments” of the dependent variables we care about (Steiner, Atzmüller and Su, 2016, 53).

Political scientists seeking to navigate these tradeoffs are usually exposed to one or the other of these schools of thought regarding experimental design, but have relatively little systematic evidence about how to choose between them.¹ Some scholars advise that respondents perform better in more concrete and familiar settings (Reiley, 2015), while others worry that doing so reduces experimental control (Camerer, 1997). Further, it has been our experience that this issue is com-

¹How political scientists do navigate these tradeoffs may have more to do with institutional path dependence rather than verifiable data or theoretical frameworks.

pounded in the publication process, where reviewers often have very strong priors about how to best navigate these tradeoffs.

And yet, as a discipline we know relatively little about the tradeoffs inherent in abstract versus concrete experimental designs. Certainly, increasing “color in the laboratory” *may* trigger “unknown (to the experimenter) impressions and memories of past experiences over which the experimenter has no control” (Friedman, Friedman and Sunder, 1994), but it is not obvious why sparse experiments would fare better in this respect. In fact, a review of the broader experimental literature suggests strong disagreement on which would be a bigger problem in terms of respondents “filling in the blanks”: rich, detailed experiments (e.g., Friedman, Friedman and Sunder, 1994) or abstract, sparse studies (e.g., Alekseev, Charness and Gneezy, 2017). And while others have noted that there is no “general theory that would give experimentalists guidance as to when stylization” might pose problems (Dickson, 2011, 61), and that this is “ultimately, an empirical issue that would have to be thrashed out by comparing data from abstract as well as contextually rich experiments” (Friedman, Friedman and Sunder, 1994, 53-4), there is surprisingly little systematic work that does so, forcing experimentalists in political science to rely on hunches and intuitions rather than systematic evidence and theoretical guidance.

In this article, we seek to make both a theoretical and an empirical contribution. First, we offer an overarching conceptual framework outlining three different dimensions of abstraction implicated in experimental design: *situational hypotheticality*, *actor identity*, and *contextual detail*. We argue that there are certain types of questions where ethical or feasibility considerations mandate at least some form of hypotheticality or abstraction, while there are others where scholars have more leeway. Yet, in those cases where scholars do have leeway, we argue that the tradeoffs between abstraction and detail in experimental design are not as stark as political scientists often claim. Second, like other recent work seeking to subject conventional wisdom about experimental design principles to empirical scrutiny (Mullinix et al., 2015; White et al., 2018; Coppock, 2019; Lupton, 2019; Mummolo and Peterson, 2019; Kertzer, 2020), we test our theoretical framework, replicating three popular vignette-based survey experiments in political science, and manipulating their levels of abstraction in three different ways. We find no evidence that situational hypotheticality changes the results experimenters obtain, an important finding as Institutional Review Boards (IRBs) and our field more broadly become increasingly concerned about the use of deception. Whether with politicians in American politics experiments, or countries in International Relations experiments,

we generally find little evidence that varying the abstraction of actor identities changes experimental results, although cue-taking experiments that use real and highly salient cuegivers obtain stronger effects than those that use fake ones. And, we show that adding contextual detail to experimental vignettes attenuates the size of treatment effects because respondents are less likely to be able to recall the treatment, but doesn't change the substantive conclusions reached. Altogether our results suggest that there are fewer tradeoffs between abstraction and detail in survey experiment design than political scientists often assume.

1 *Abstraction and detail*

One of the many design choices political scientists face when using experimental methods concerns the appropriate level of *abstraction*. There is a rich literature on abstraction in philosophy, psychology, and cognitive science, which often operationalizes abstraction in slightly different ways (e.g., [Cantor and Mischel, 1979](#); [Paivio, 1990](#); [Semin and Fiedler, 1988](#); [Colburn and Shute, 2007](#)). For our purposes, we borrow from construal level theory in defining abstraction as a higher-level representation ([Sartori, 1970](#), 1040-46; [Trope and Liberman, 2003](#)). It involves making “a distinction between primary, defining features, which are relatively stable and invariant, and secondary features, which may change with changes in context and hence are omitted from the higher-level representation” ([Shapira et al., 2012](#), 231). As [Table 1](#) shows, an abstract representation is sparse and decontextualized, reduced to the object's most central elements (e.g., “A nuclear weapon”), whereas a concrete representation is contextualized and rich in specific detail, including subordinate considerations (e.g., “North Korea's Hwasong-14 intercontinental ballistic missile”).

Table 1: Conceptualizing abstraction

Abstract	Concrete
High level of construal	Low level of construal
Decontextualized	Contextualized
Primary features	Secondary features
Sparse	Complex
Superordinate elements	Subordinate elements

Modified from [Trope and Liberman \(2003, 405\)](#).

Two points here are relevant for our purposes. First, in specifying which elements of a construct are primary and which are secondary, the act of abstraction is inherently a theoretical phenomenon,

rather than an empirical one. Although our interest in this article is in abstraction in experimental methods, similar questions also arise in formal modeling, and in quantitative and qualitative methods alike (Clarke and Primo, 2012).

Second, questions about the appropriate level of abstraction loom large in a variety of issues in experimental design: whether experiments should be “stylized” or “contextually rich” (Dickson, 2011; Kreps and Roblin, 2019), use real or hypothetical actors (McDonald, 2019; Nielson, Hyde and Kelley, 2019), and refer to imminent, future, or hypothetical situations. In this sense, experiments can be abstract or concrete along multiple dimensions at the same time. In the discussion below, we suggest that abstraction in experimental design can be conceptualized along at least three dimensions: situational hypotheticality, actor identity, and contextual detail. We classify a set of recent experiments in political science along these dimensions in Table 2, and review each dimension in detail in the discussion below.

Table 2: Abstraction in experimental political science

Type of experiment	Example	Type of abstraction		
		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
Vignette-based 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

1.1 SITUATIONAL HYPOTHETICALITY

The first type of abstraction in experimental design concerns whether a scenario is described as hypothetical or not.² The rationale for using hypothetical scenarios in survey experiments is simple: in their most stylized form, experimentalists make causal inferences by drawing comparisons between two different states of the world, randomly assigning participants to either a treatment condition, or control. Some experiments intervene by giving respondents in the treatment condition information about the world that they might not otherwise have (e.g., Butler, Nickerson et al.,

²Note that the relevant question here is whether the scenario is *presented* as hypothetical, rather than whether the details described in the scenario are in fact hypothetical; as we note below, experiments utilizing deception, for example, may present hypothetical scenarios as if they are real.

2011; Raffler, 2019), but especially in survey experiments, experimentalists often manipulate features of the world itself. In order to manipulate features of the world in this manner, experimentalists must either engage in deception (showing respondents mock news articles purported to be real, e.g., Brader, Valentino and Suhay, 2008; Arceneaux, 2012), or find another way to justify—whether to respondents, or to Institutional Review Boards (IRBs)—why the scenario being described to respondents deviates from the one they are in.

There are a variety of techniques experimentalists often employ. One is to explicitly describe the scenario as hypothetical: respondents in Boettcher (2004, 344), for example, are asked to “envision a hypothetical presidency apart from the current administration.” Others implicitly invoke hypotheticality: respondents participating in conjoint experiments studying immigration preferences, for example (e.g., Hainmueller and Hopkins, 2015), are presumably not under the illusion that the immigrants they are being asked to choose between are real. Another widely used variant under the category of “implicit hypotheticality” is to describe a scenario as set in the future. Mattes and Weeks (2019), for example, tell respondents “We are going to describe a situation the U.S. could face in the future, in 2027.” This is often termed a *prospective* scenario, but ultimately the future setting is simply a mechanism to make the scenario implicitly hypothetical.

The rationale for these design choices are often not explicitly stated, but usually involve concerns that respondents will not take studies as seriously when scenarios are presented as explicitly hypothetical — the sense that researchers asking hypothetical questions will be rewarded with hypothetical answers (Converse and Presser, 1986, 23). Experimentalists operating out of an economics-style tradition tend to avoid both deception and situational hypotheticality in order to accentuate the effects of incentives (Morton and Williams, 2010). Yet, there is relatively little empirical work testing the conditions in which situational hypotheticality affects responses in political science experiments.

1.2 ACTOR IDENTITY

The second dimension of abstraction involves the identity of the actors invoked in experimental vignettes: are they real, or artificial? Some experimenters explicitly use real world actors in contexts ripped from the headlines, as in Boettcher and Cobb’s (2006) study of how casualty frames shape support for the war in Iraq, or Evers, Fisher and Schaaf (2019), who experimentally investigate audience costs using Donald Trump and Barack Obama. In this sense, the artificiality of the actors

in an experiment is distinct from the hypotheticality of the situations in which actors are embedded. Indeed, experimenters often use real world actors in hypothetical scenarios. For example, Kriner and Shen's (2014) casualty sensitivity experiments explore how many casualties Americans would be willing to bear in a series of "hypothetical" interventions in "real" countries (Somalia, Darfur, Iran, and Liberia). In this case, the military interventions are artificial and prospective, while the relevant target countries are real.

Moving up the ladder of abstraction, some experimenters describe hypothetical scenarios in artificial countries, in order to exert complete control over how much information participants bring to bear. For example, Brooks and Valentino (2011) describe a conflict between "Malaguay and Westria", and Rubenzer and Redd (2010) describe a crisis in the state of "Gorendy." Taking this approach a step forward, many experimentalists use unnamed countries, describing target states as "Country A" or "Country B" (Johns and Davies, 2012; Yarhi-Milo, Kertzer and Renshon, 2018), or simply referring to "A country" rather than providing a label (Tomz and Weeks, 2013).³

Concerns about actor identity and hypotheticality are not limited to the subfield of international relations. In comparative politics, Banerjee et al. (2014) describe hypothetical representatives (running for office in hypothetical districts) to study the concerns of voters in rural India. "Hypothetical candidate" experiments are also a long-running feature in the study of American politics (as in Rosenwasser et al., 1987; Colleau et al., 1990; Kam and Zechmeister, 2013) — and are particularly common in conjoint experiments — although the results here are mixed. In a meta-analysis of 111 studies of negative campaigning, Lau, Sigelman and Rovner (2007) find that experiments featuring hypothetical candidates don't offer significantly different results from those featuring real ones. McDonald (2019), in contrast, argues that experiments on hypothetical candidates both increase cognitive burden and produce larger treatment effects than experiments on candidates about which respondents have strong priors.

As with the case of situational hypotheticality, the logic of using unnamed or hypothetical actors stems directly from the questions being tested. Political scientists turned to experimental methods to study the effects of candidate gender (Huddy and Terkildsen, 1993; Sanbonmatsu, 2002; Brooks and Valentino, 2011), for example, precisely because it is difficult to find two real-world candidates identical to one another on all dimensions other than their gender. The same is true in studies of race in politics (Burge, Wamble and Cuomo, 2020; Wamble, 2020), or ethnicity

³In instructions to participants, researchers often note that this generality is "for scientific validity."

([Dunning and Harrison, 2010](#), though see [Adida, 2015](#)). In an IR context, it is hard to think of two real-world countries that are identical in all respects but one, such that IR scholars interested in manipulating the effects of regime type, military capabilities, or foreign policy interests usually do so with fictional or hypothetical countries (e.g., [Rousseau and Garcia-Retamero, 2007](#)).

1.3 CONTEXTUAL DETAIL

The third dimension of abstraction involves the amount of additional context provided in an experiment. [Press, Sagan and Valentino \(2013\)](#) present a lengthy newspaper article that provides participants with a large amount of context, as do experiments in American politics that generate fake campaign advertisements or news clips ([Brader, Valentino and Suhay, 2008](#)). In contrast, other experiments often present relatively little information. [Trager and Vavreck \(2011\)](#), for example, manipulate the President's strategy in a foreign policy crisis as well as information about the US domestic political environment, but as with most audience cost experiments, they say relatively little about the context of the intervention itself. Most experiments in the experimental economics tradition offer relatively little contextual detail at all, deliberately paring the design down to be void of explicit political content (e.g., [Tingley and Walter, 2011b](#); [Kanthak and Woon, 2015](#); [Quek, 2017](#)).

The argument usually offered in favor of contextual detail is that it increases realism and respondent engagement. Anecdotally, in some of our own research, when we include open-ended prompts at the ends of survey experiments soliciting feedback from participants, one of the most frequent comments we have received involved scenarios being "too vague": participants wanted more information in order to help them make up their minds. Yet apart from [Kreps and Roblin \(2019\)](#) and [Bansak et al. \(2020\)](#), there has been little empirical work to adjudicate what the consequences of providing richer or sparser stimuli might be. [Bansak et al. \(2020\)](#) use a clever multi-stage conjoint design to first find "filler attributes" (information uncorrelated with the object of interest in the study) and then experimentally vary the amount of filler in the second stage, finding relatively stable treatment effects even with large numbers (up to 35) of filler items. [Kreps and Roblin \(2019\)](#) focus on treatment "formats," in particular the difference between information presented in mock news stories versus short/long vignettes (with a slight conflation of the "length" versus "format" dimensions), finding that respondent attention (as a measure of satisficing) was unaffected by the presentational format.

This discussion suggests that what is often referred to as “contextual detail” is actually composed of at least three related dimensions. The first is simply the volume of information provided: more or less information can be provided in an experiment to supplement the treatments and add “realism.” We might, for example, provide a long or short biography of a candidate or background to an international dispute. The second concerns *how* the information is presented, and here there have been examples of any number of treatment formats in experiments, from bullet-pointed vignettes (Tomz, 2007), to mock news stories (Druckman and Nelson, 2003) and “ad-watch” style reports (Valentino, Neuner and Vandebroek, 2018). The third is the content of the information itself, which is orthogonal to its volume. Any bit of information may be classified as either what Bansak et al. (2020) call “filler” or its opposite, what we term “charged” content, which may interact with the treatment in some way and affect the results of a study through a mechanism other than simple respondent satisficing. If a President’s “favorite highway” is filler, then Bansak et al. (2020) also show that other attributes (e.g., previous occupation and number of children) are associated with the object of interest and are thus ill-suited to be added simply to increase the “realism” of a vignette. But while they show that satisficing is less of a problem than we might expect once we introduce filler attributes, we are still largely in the dark with respect to understanding how the addition of charged (versus filler) content affects our interpretation of experimental results.

2 *Navigating the tradeoffs*

In sum, although political scientists tend to recognize that tradeoffs between abstract and concrete experiments exist, there is less certainty about how one should balance them. Often, for example, political scientists run both abstract and concrete versions of their experiment to test whether the results hold (e.g., Herrmann, Tetlock and Visser, 1999; Levine and Palfrey, 2007; Rousseau and Garcia-Retamero, 2007; Berinsky, 2009; Horowitz and Levendusky, 2011; LeVeck et al., 2014; Renshon, Dafoe and Huth, 2018; Nielson, Hyde and Kelley, 2019). However, doing so can be somewhat inefficient, as adjusting levels of abstraction on multiple dimensions simultaneously provides limited insight regarding the specific dimension driving experimental outcomes.⁴

There are some circumstances where for logistical or ethical reasons, experimenters will be constrained in terms of how abstract or concrete their stimuli will be. For example, researchers are

⁴Moreover, given the presence of a budget constraint, running two versions of the same experiment to preempt questions about abstraction can also lead to statistical power problems.

limited in their ability to select real world actors when studying the effects of race and gender in candidate selection, or the effects of country-level characteristics on foreign policy preferences. Additionally, there are experiments where some form of situational hypotheticality is required (often at the demand of IRBs) to avoid the use of deception, and some contexts where the use of deception raises ethical challenges: for example, telling respondents that a real-world political candidate is unethical (e.g., [Butler and Powell, 2014](#)). Experimentalists using subject pools that prohibit deception, or seeking to publish in journals that do the same ([Morton and Williams, 2010](#)), will face similar restrictions in choosing the optimal level of abstraction in experimental design.

In other cases, however, experimentalists have more of a choice when designing their studies. In the discussion below, we link each dimension of abstraction to questions about experimental control, on the one hand, and generalizability, on the other. Although political scientists often see these two principles as in tension with one another — associating the former with internal validity, and the latter with external validity — we argue that the implications of abstraction in experimental design for each principle are actually more complex. There are some instances where an increase in abstraction may enhance experimental control, and others where an increase in abstraction may come at the expense of experimental control; because experimentalists may not exercise as much control over their respondents as we like to think, more abstract stimuli may not necessarily be more generalizable. We suggest, then, that the tradeoff between abstract and concrete experimental designs represents something of a paradox: the circumstances in which experimentalists have the most leeway in terms of the abstraction of design choices may be the ones where the tradeoffs between different design choices are the least consequential.

2.1 EXPERIMENTAL CONTROL

Experimenters seek to obtain “control” over the ways in which respondents construe the contextual features of vignettes, in order to ensure proper implementation of their experimental designs.⁵ When experimental vignettes invoke different reactions amongst different types of respondents — or perhaps invoke reactions the researcher never intended — experimenters can risk losing control over their study, raising concerns regarding internal validity. By varying the information provided

⁵Similarly, experimenters seek to control the construal of treatments themselves. As [Tomz and Weeks \(2013\)](#) point out, if participants who are told that a country is democratic are also likely to assume that democracies are more likely to win the wars they fight, the regime type treatment becomes double-barreled, manipulating omitted details (see also [Kertzer and Brutger, 2016](#); [Dafoe, Zhang and Caughey, 2018](#)).

along the three aforementioned levels of abstraction, experimenters can potentially shape the degree of control they obtain.

Yet we argue that there are less to these tradeoffs than meets the eye. First, the relationship between abstraction and control varies based upon the dimension under investigation. Increasing contextual detail is often thought to enhance experimental control, by fixing the type and degree of information that all subjects share regarding an issue area. For example, when implementing an endorsement experiment regarding a (fictional or real) immigration policy (Nicholson, 2012), researchers can provide detailed information regarding: i) who initiated the policy, ii) when it comes into effect, and iii) how it relates to previous policies. Presumably, this information can ensure an informational common denominator, and avoid a situation in which respondents with different background knowledge construe the experimental vignette in diverging ways.

In contrast, increased detail in terms of actor identity is usually argued to reduce experimental control. In an international relations context, Herrmann, Tetlock and Visser (1999, 556) note that “the use of real countries [adds] a degree of realism...but it also sacrifice[s] a degree of experimental control. Affective reactions to the various countries may differ, and [characteristics of the countries] may not be perceived uniformly by all participants.” In American politics, Reeves and Rogowski (2018, 428) write that “the use of hypothetical candidates comes at the cost of reducing the real-world attributes of the experiment, but this cost is offset by removing respondents from their feelings about any actual politician, which could serve as confounders.” These examples suggest that by introducing real world actors and adding detail into vignettes, experimenters lose control over their respondents — the opposite of conventional wisdom about the effects of contextual detail.

More generally, it may be somewhat misleading to think that by turning from real to hypothetical actors, or from contextually sparse to rich vignettes, experimenters gain control over their study. Indeed, when presented with relatively pared down stimuli, participants often “fill in the blanks.” For example, scenarios in which “a country sent its military to take over a neighboring country” in which the US is considering sending troops to repel the invader (Herrmann, Tetlock and Visser, 1999; Tomz, 2007; Trager and Vavreck, 2011), may lead participants to think of the Gulf War.⁶ More-

⁶Similarly, in their survey experiment on the effect of regime type on support for the use of force, Johns and Davies (2012, 1043) note that the vignette is “loosely based on recent debates about Iran but [makes] no explicit reference to that or any other real country,” acknowledging the possibility that at least politically sophisticated participants are likely thinking of Iran when they answer the questions.

over, different types of respondents may invoke different mental schema when faced with the same abstract scenario; depending on a respondent's generation, "repel an invader" experiments may invoke Korea, Vietnam, Iraq, or none of the above. Likewise, it is possible that different respondents will exert diverging reactions to additional contextual detail, leading experimenters to lose, rather than gain control. Adopting an abstract design can thus both increase or decrease experimental control, such that the tradeoff here may not be as clean cut as experimentalists sometimes suggest.

Even if experimenters may have more leeway when choosing the appropriate level of abstraction for actor identity than is often claimed, this does not mean that all concrete actor identities are equally desirable. In particular, experimenters should attend to at least two considerations when choosing real world actors. The first is *schema consistency* (Hashtroudi et al., 1984): is the choice of actor reasonable given the scenario in which the actor is embedded? For example, in experimental scenarios in which a country is pursuing a nuclear weapons program (e.g., Tomz and Weeks, 2013), experimental control decreases if the experimenter chooses a country that already has nuclear weapons (e.g., Russia), or a country that respondents think is unlikely to pursue them (e.g., Canada). If a schema-inconsistent actor is chosen, the respondent is less likely to believe the scenario or accept the treatment, thus weakening the treatment effect. The second is *treatment consistency*: if the treatment manipulates an attribute of an actor, are all of the levels of the attribute being manipulated seen as plausible by respondents? In candidate selection experiments, for example, it would be difficult to manipulate the partisanship of politicians with well-known partisan identities, or to manipulate the policy stances of politicians on issues where they have already taken prominent positions. If respondents do not perceive the treatment as consistent with the identity of the actor, then the experimenter is likely to lose control since the respondent may not comply with the treatment, attenuating the treatment effect.

2.2 GENERALIZABILITY

While experimental control is a fundamental aspect in designing vignettes, scholars may very well be concerned by other factors such as generalizability – the extent to which results from a given study speak to a broader set of real world scenarios. Like control, degrees of generalizability may be shaped by levels of abstraction in experimental design. Thus when framing an experiment as hypothetical or real, and when selecting particular actors, and levels of contextual detail, researchers may condition the degree to which their results generalize beyond a particular context.

Oftentimes, experimenters adopt unnamed actors in experimental vignettes in order to enhance generalizability. At least implicitly, the selection of an unnamed actor is motivated by the fact that a researcher's quantity of interest is a main effect rather than a conditional effect. In other words – for example, the researcher is interested in the effect of past behavior on forming reputations for resolve in general, not the effect of past behavior on forming reputations for resolve for Iran specifically (Renshon, Dafoe and Huth, 2018).

Yet it is unclear that increased abstraction actually increases generalizability. First, when we generalize from these experiments to the problems in the real world that motivate us to conduct them in the first place, selecting unnamed actors may lead us to miss important sources of treatment heterogeneity, and may even make it harder to generalize results to any motivating real world cases.⁷ For example, because respondents are often “pre-treated” with partisan cues prior to participating in our studies (Gaines, Kuklinski and Quirk, 2007), experimenters might deliberately choose nonpartisan scenarios where these pretreatment effects are minimized, lest the effects of partisanship swamp or overwhelm the treatments of interest. Yet if many political phenomena have a partisan hue, the absence of partisan dynamics in the experiment actually makes it harder to generalize these results (McDonald, 2019).

Similarly, the degree of contextual detail provided by experimenters might shape the extent that findings from an experiment can generalize to real world scenarios. If participants in experiments only receive two pieces of information, one of which is the treatment being randomly assigned, the relative “dosage” of the treatment is likely to be unrealistically high, and may not hold in a more naturalistic setting (Barabas and Jerit, 2010). In contrast, if the treatment is presented to participants embedded in a larger amount of information (a full newspaper article, rather than just a few bullet points, for example), the treatment is likely to exert a (realistically) smaller effect. The same consideration holds with situational hypotheticality: rather than bolster generalizability, it may in fact decrease it, if when given strictly hypothetical scenarios, respondents provide strictly hypothetical answers (Converse and Presser, 1986).

In sum, then, although experimentalists frequently think about questions regarding experimental control and generalizability as two competing principles, the latter linked to abstract designs,

⁷This is true both for contextual treatment heterogeneity in which treatment effects vary with features of the situation not being modeled in the study (e.g., do the treatment effects in a relatively abstract scenario reflect the treatment effect for an “average” country?), or in respondent-level treatment heterogeneity, in which the treatment effects in the real world vary with characteristics of respondents that would not manifest themselves in a highly stylized scenario.

and the former to concrete ones, it is not clear that the tradeoffs are actually as stark: adding contextual detail can increase control, but choosing real-world actors may lower it; respondents given pared down stimuli can still “fill in the blanks”, such that more abstract designs are not necessarily more generalizable than concrete ones. We seek to evaluate these conjectures empirically. Specifically, by experimentally manipulating the situational hypotheticality, actor identity, and contextual detail of a series of popular experiments, we aim to determine if and how different forms of abstraction shape the results of experiments. If introducing real actors or elaborate contextual detail systematically affects experimental control and generalizability, then one would expect to observe variation in outcomes across experiments varying in abstraction. If, however, the amount and type of detail across experiments only modestly shapes the results, then it would appear that the tradeoffs between these design choices are somewhat overstated.

3 Research Design

To provide guidance for experimentalists on how abstract their experiment *ought* to be as well as how scholars should balance the potential tradeoffs associated with differing levels of abstraction, we fielded a series of survey experiments across two separate studies, each designed to address one of the dimensions of abstraction described earlier. We do so partly by building on three recent vignette-based survey experiments in political science: Nicholson’s (2012) study of elite cues in American immigration policy debates (which we refer to as the “Elite Cue” experiment), Press, Sagan and Valentino’s (2013) test of the nuclear taboo in public opinion towards the use of force (the “Nuclear Weapons” experiment), and Mutz and Kim’s (2017) exploration of the role of in-group favoritism in trade preferences (the “In-group Favoritism” experiment). The nuclear weapons and in-group favoritism experiments were fielded in a single survey, with the order of the experiments randomized, while the elite cues experiment was fielded in a separate survey.

As depicted in Table 3, our study is comprised of three levels of treatment: (1) the central treatments from the replicated studies, (2) contextual detail and actor identity treatments varying the amount of context or the names of the actors respondents are presented with, and (3) a situational hypotheticality treatment which describes experimental scenarios as either real, explicitly hypothetical, or implicitly hypothetical.⁸

Our selection of three different studies allows us to devote considerable attention to the ways

⁸An additional summary of the structure of our survey instrument is depicted in Appendix §1.

in which diverse types of abstraction and concreteness, relating to situational hypotheticality, actor identity, and contextual detail, affect the inferences drawn from experimental designs. The elite cues experiment replicates Nicholson's (2012) study of elite cues about immigration reform in the United States, to explore the effects of actor identity in experimental design.⁹ The original study assigned respondents to read about an immigration proposal endorsed by one of two highly salient partisan elites or not endorsed by any actor. In our replication, we added actor identity treatments that vary whether the immigration reform endorsement is made by less salient partisan cuegivers, or by a fictional politician whose partisanship we manipulate. These treatments allow us to test how moving from more abstract (whether fictional, or low salience) actors to more concrete ones (high-salience and real) actors modifies experimental findings.

The ingroup favoritism experiment replicates portions of Mutz and Kim's (2017) investigation of American trade preferences to study the effects of additional contextual detail. We build on their design 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. (2020), 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.

The nuclear weapons experiment replicates 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. First, we manipulate the identity of the country in which the potential attack would take place, choosing either an unnamed country, a fictional country, or a country that is either schema-consistent with the setup of the experiment, or schema-inconsistent. 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 Press, Sagan and Valentino (2013) in its un-named country format. We then presented the study's main

⁹While 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).

outcome questions, and asked respondents to rate the likelihood that each of eleven different countries would be the actor in each scenario.¹⁰ Throughout the discussion below, we refer to countries rated as the most applicable for an experimental scenario as *schema-consistent*, and countries rated the least applicable as *schema inconsistent*, thereby ensuring our selection of specific country names depicted in Table 3 is empirically grounded.

Lastly, throughout all of the studies we introduce a situational hypotheticality treatment (randomized at the subject-, not the study level) which refers to the depicted scenarios as either real, explicitly or implicitly hypothetical in order to test whether manipulating hypotheticality moderates the experimental findings.¹¹ The details of the three individual studies are below, and depicted in Table 3. The in-group favoritism and nuclear weapons experiments were fielded on a sample of $N = 4686$ respondents through Dynata in spring 2019. The elite cues experiment was fielded on a sample of $N = 4070$ respondents through Lucid’s “Theorem” respondent pool in spring 2020.¹² Participants in each survey were informed that they were participating in a study, were asked to complete a voluntary informed consent form, and were compensated for their participation based on the practices of the survey firms.¹³

3.1 ELITE CUES EXPERIMENT: MANIPULATING ACTOR IDENTITY

In our extended replication of Nicholson’s (2012) study we focus on a common dilemma for experimental political scientists: whether to name specific actors — whether politicians, countries or organizations — in a vignette or leave them unspecified, and if the actors are named, whether they should be highly recognizable, relatively low-salience, or made up from scratch. 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 replication, 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

¹⁰For more information regarding our pretest procedure see Appendix §3.

¹¹In our first survey respondents were assigned to one of two conditions describing a situation as either implicitly or explicitly hypothetical. In our second survey respondents were assigned to one of three conditions describing a situation as either real, implicitly, or explicitly hypothetical.

¹²More details about each platform are available in Appendix §1

¹³For a broader discussion of the ethical considerations and practices of the research design, see Appendix §1.

	Elite Cues: Nicholson 2012	In-Group Favoritism: Mutz & Kim 2017	Nuclear Weapons: Press, Sagan & Valentino 2013
Treatments from original study	1. No Endorsement 2. In-Party Cue 3. Out-Party Cue	1. US gains 1000 and other country gains 10 2. US gains 10 and other country gains 1000 3. US gains 10 and other country loses 1000	1. 45% Success for conventional attack 2. 90% Success for conventional attack
Actor identity and contextual detail treatments	If assigned to cue: 1. Real + High Saliency (Donald Trump/Joe Biden) 2. Real + Low Saliency (Mike Rounds/Tom Carper) 3. Fictional (Stephen Smith/Stephen Smith)	1. No Additional Context (original) 2. Filler Context 3. Charged Context	1. Extended Context (original) 2. Reduced context 1. Unnamed (foreign country) 2. Made up (Malaguay) 3. Real + Schema consistent (Syria) 4. Real + Schema inconsistent (Bolivia)
Situational hypotheticality treatment	Situation described as: 1. Implicitly hypothetical 2. Explicitly hypothetical 3. Real	Situation described as: 1. Implicitly hypothetical 2. Explicitly hypothetical	Situation described as: 1. Implicitly hypothetical 2. Explicitly hypothetical
Factorial design: (total cells in parentheses)	$(1 \times 3) + (2 \times 3 \times 3) = 21$	$3 \times 3 \times 2 = 18$	$2 \times 2 \times 4 \times 2 = 32$

Table 3: Summary of Treatments for 3 Studies

fictional politician (Stephen Smith) whose partisanship we manipulate.¹⁴

Our procedure for the [Nicholson](#) replication is detailed in Appendix §2.1. Like [Nicholson](#) we present respondents with an experimental scenario in which information is presented about an immigration reform policy “in the news.” Depending on which treatment arm they were assigned to, respondents also read that the proposal was backed by either Donald Trump, Joe Biden, Mike Rounds, Tom Carper, Stephen Smith or did not receive any additional information about endorsements. 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 particular 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.

¹⁴Additionally, we update the substantive context of the experiment to focus on protection for “Dreamers” in the U.S.

3.2 INGROUP FAVORITISM EXPERIMENT: MANIPULATING CONTEXTUAL DETAIL

Mutz and Kim (2017) use a survey experiment to examine the impact of in-group favoritism on American attitudes towards international trade, by randomizing the expected US gains and losses from a trade agreement, in comparison to another country. In replicating their basic framework, we focus on a common decision experimentalist grapple with when designing instruments: how much contextual detail should vignettes include? Should they be stripped-down, bare-bones vignettes that highlight salient factors, or context-rich stories that in their level of detail more closely resemble the volume of information we might encounter in our daily lives?

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).

A detailed description of our replication protocol is provided in Appendix §2.2. In a similar fashion to our replication of Nicholson, 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.

3.3 NUCLEAR WEAPONS EXPERIMENT: MANIPULATING BOTH CONTEXTUAL DETAIL AND ACTOR IDENTITY

In our final study, we replicate Press, Sagan and Valentino's (2013) experiment on "atomic aversion." They 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. They do so by randomizing the relative probability of success for conventional attacks relative to nuclear attacks.¹⁵ 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

¹⁵The original study includes three treatment conditions where conventional attack 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.

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).

As detailed in Appendix §2.3, apart from our two additional treatment arms and the simplification of original treatments, we follow a similar procedure to [Press, Sagan and Valentino](#). After the main outcome measure, we present respondents with a thought listing exercise and factual questions relating to the main treatment.

4 Results

4.1 REPLICATION OF ORIGINAL STUDY RESULTS

In [Figure 1](#) we present our initial replication of the three studies under investigation. In the lower portion of the figure, we plot the main treatment on our replication of Mutz and Kim’s study of trade and in-group preferences. More specifically, the treatment in this study is exposure to a vignette describing a trade deal in which the U.S. gains 1000 jobs, in relation to a trading partner who loses 10 jobs. Respondents in the comparison conditions were either told that (i) the U.S. will gain 10 jobs, in relation to a trading partner who will gain 1000 jobs, or that (ii) the U.S. will gain 10 jobs, in relation to a trading partner who will lose 1000 jobs. Our treatment essentially exposes respondents to a trade deal that highly benefits (treatment) or doesn’t really benefit (control) the U.S., and as demonstrated in [Figure 1](#), respondents in the treatment condition are much more likely to support trade deals when the US is expected to gain more.¹⁶

In the middle region of [Figure 1](#), we present our replication of Nicholson’s immigration-policy endorsement experiment. Specifically, we consider how exposing partisans to out-party cues supporting a given immigration reform affects support for the named reform. As evident from the study’s original data (depicted in [red](#)), as well as our replication (depicted in [blue](#)), out-party endorsement cues have a negative causal effect on support for immigration policy.¹⁷ Notably effect sizes are comparable in the original study and our replication, and standard errors are smaller in the replication, likely as a result of our substantially larger sample size.

Lastly, in the upper portion, we present our replication of the Press, Sagan, and Valentino study.

¹⁶We do not include the original data estimate for Mutz and Kim because the original study included a more complex design with the potential for each country to gain or lose 1, 10, 100, and 1000 jobs, in contrast to our simplified version.

¹⁷For ease of interpretation, in these analyses we compare respondents receiving in-party cues with respondents receiving out-party cues, omitting all respondents in the no-cue condition.

Figure 1: Replication of ATEs from the three experiments

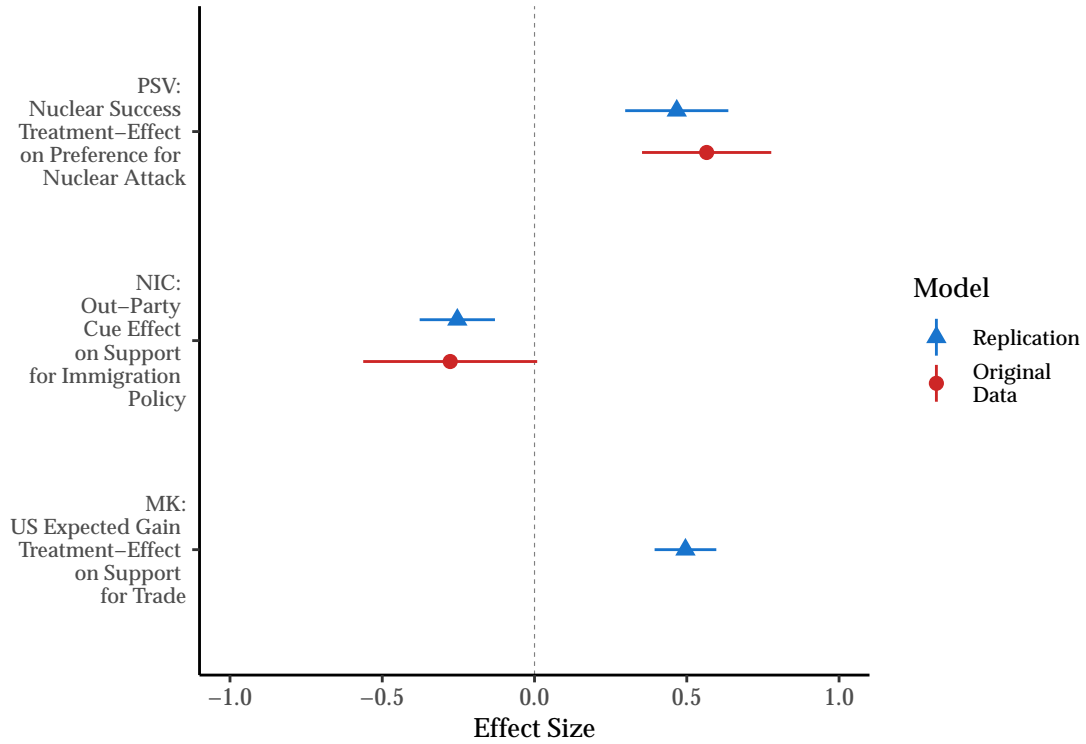


Figure 1 shows we successfully replicate the average treatment effects from the original studies. Point estimates and corresponding confidence interval are extracted from separate OLS models where original outcomes are predicted by treatments. When possible we benchmark our replication (Blue) to original studies (Red). In doing so we consider respondents exposed to the original format of the experiment, omitting respondents exposed to new variants of the experiment where we introduce diverging elements of abstraction or detail. All outcomes are standardized.

Doing so, we test whether exposing respondents to a vignette in which a nuclear attack is more effective than a conventional attack, alters support rates for nuclear attacks. As evident, we replicate Press, Sagan, and Valentino’s initial results, demonstrating that respondents are more likely to support a nuclear attack when it is described as more effective than a conventional attack. Our replication of average treatment effects (depicted in blue) is substantively similar to replications using the original data (depicted in red).

Taken together, the results in Figure 1 demonstrate our initial success in replicating our studies of interest. Now we turn to consider how our additional treatments moderate the main results depicted above. We start by considering our hypotheticality treatment. Then, we turn to consider the moderating effects of actor identity and contextual detail.

4.2 SITUATIONAL HYPOTHETICALITY EFFECTS

Does describing an experimental scenario as explicitly hypothetical, prospective, or real moderate identified effects in experimental designs? To answer this question, we administered our situational hypotheticality treatment which assigned respondents to introductions describing each experimental vignette in a particular way. Specifically, in our initial replication of Press, Sagan and Valentino and Mutz and Kim we described experimental vignettes as either explicitly hypothetical or prospective, and thus implicitly hypothetical. When considering Nicholson’s experiment in our second study, we introduced experimental vignettes as either hypothetical, real, or without addressing hypothetically at all. Throughout all our studies, subjects were randomly assigned to one hypotheticality condition at the beginning of the survey instrument. So any given subject would consistently see the same hypotheticality treatment at the start of each of our replicated experiment.

To examine the effect of this design choice, we use standard OLS models in which we interact the original treatment from a given study —e.g., in the elite cues experiment, whether an out-party politician is the endorser of the immigration reform policy — with our hypotheticality treatment. Figure 2 presents results in which our main quantity of interest is the interaction effect, representing the moderating effect of our hypotheticality treatment on the original treatments. In our elite cues replication, hypotheticality can take one of three values (explicitly hypothetical, implicitly hypothetical, or real). However, we focus on comparing the real and explicitly hypothetical conditions, which are most distinct.¹⁸

¹⁸Comparisons of the explicit and implicit hypothetical conditions yield similar results.

As evident in Figure 2, framing an experimental vignette as explicitly hypothetical does not seem to change the main findings from experimental studies. Indeed, in all models, our situational hypotheticality treatment, and its interaction with original treatments are statistically and substantively insignificant. We construe these results as evidence for the limited empirical consequences of design choices relating to situational hypotheticality.

Figure 2: No moderating effects of situational hypotheticality

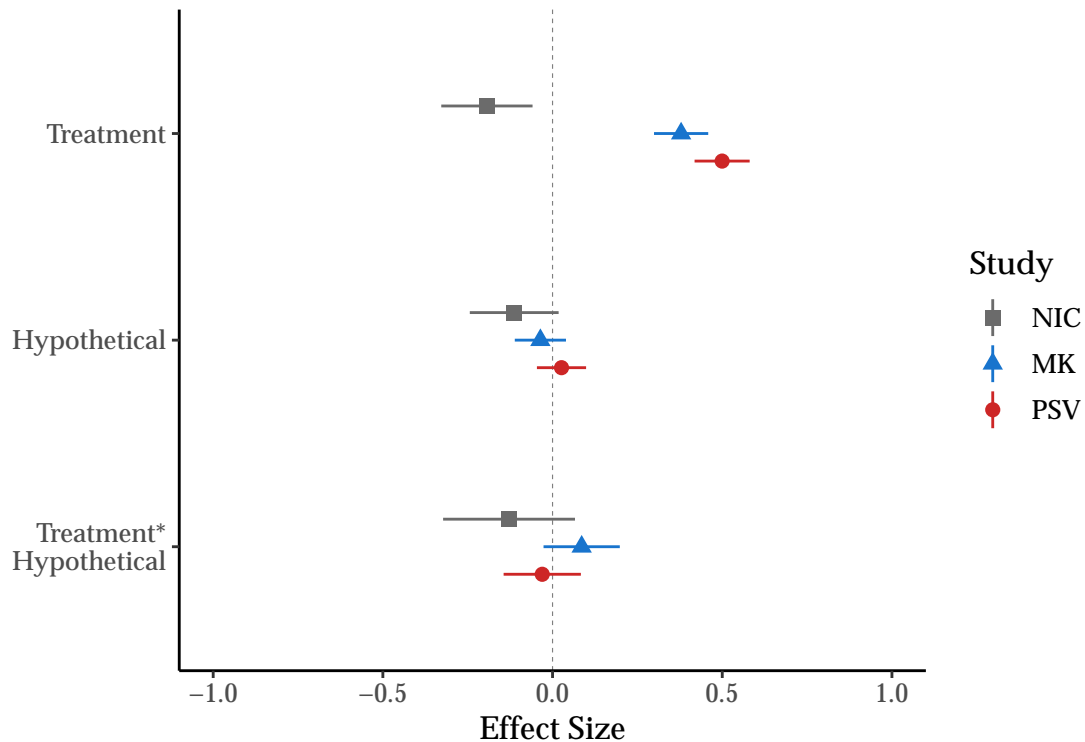


Figure 2 finds no evidence that situational hypotheticality significantly moderates our treatment effects in any of the three experiments. Point estimates and corresponding confidence intervals are extracted from three separate OLS models where original outcomes are predicted by original treatments interacted with the hypotheticality treatment. All outcomes are standardized.

4.3 ACTOR IDENTITY EFFECTS

We now turn to an analysis of how actor identities of different levels of abstraction affect findings from experimental vignettes. In our replication of Press, Sagan and Valentino’s nuclear weapons experiment we randomized the target country as: unnamed (our baseline condition), fictional (Malaguy), real and schema inconsistent (Bolivia), or real and schema consistent (Syria). Similarly,

Figure 3: Moderating effects of actor identity condition

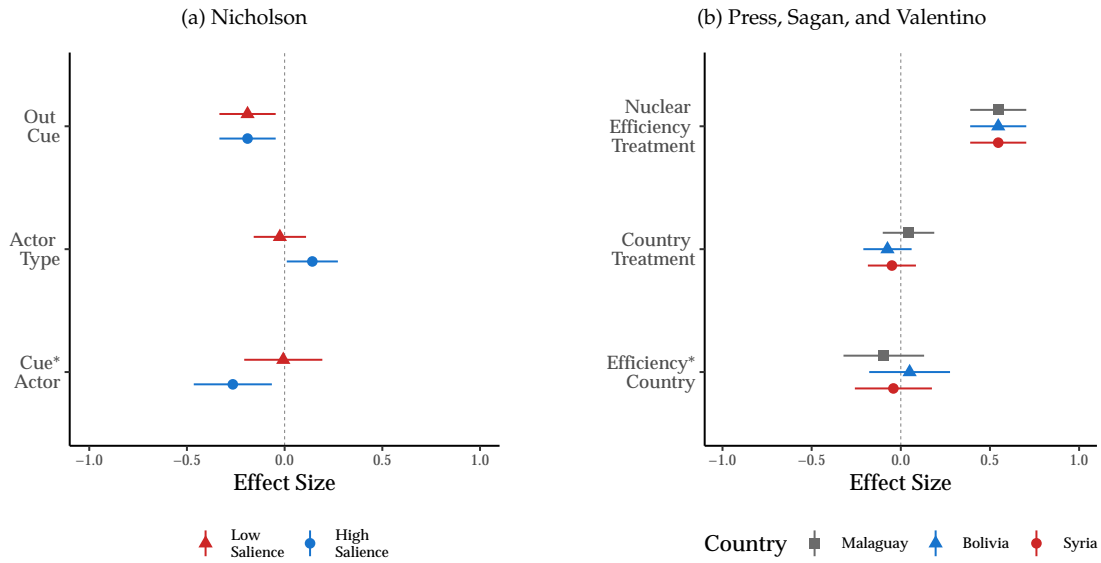


Figure 3 shows that manipulating the country identity does not significantly moderate treatment effects in the nuclear weapons experiment, although we obtain slightly larger treatment effects in the elite cue study when we use more salient cue-givers. Point estimates and corresponding confidence intervals are extracted from five separate OLS models where original outcomes are predicted by original treatments interacted with different actor identity conditions. Panel *a* compares made-up politicians with low salience (red) and high salience (blue) politicians. Panel *b* compares the unnamed country condition with a fake country name condition (gray), schema inconsistent country name condition (blue), and schema consistent country name condition (red). All outcomes are standardized.

in the Nicholson elite cues replication we randomized whether an out-party endorsement was by a made-up politician (Stephen Smith [D or R], our pooled baseline condition), a low salience politician (Senators Mike Rounds [R] or Tom Carper [D]), or a high salience politician (Donald Trump [R] or Joe Biden [D]).

We interact this actor identity treatment with each study’s original treatment, and present results for both our elite cues and nuclear weapons replications in Figure 3 (Panel A and B respectively). In these figures, our main quantity of interest is the interaction between the original treatment and our additional actor identity treatment.

As demonstrated in Figure 3, most actor identity conditions do not seem to moderate main treatment effects. It follows that for the most part, whether an actor is unnamed, fictional or real — and if real, schema-consistent or inconsistent — does not lead scholars to draw substantively different inferences or identify diverging effects, either in magnitude or direction. That said, in the

left panel of Figure 3, we show that using high salience actors amplifies the endorsement treatment effects (when compared to baseline made-up actors).

There are at least three potential mechanisms to explain the actor identity results from the elite cue experiment. The first potential mechanism is cognitive burden. McDonald (2019) proposes a version of this hypothesis, arguing that survey experiments using hypothetical actors increase the cognitive burden on respondents, as measured by response latencies in survey questions. Yet as we show in Appendix §4, there is no significant effect of the actor identity treatment on response latency in our study, so it does not appear that moving from a hypothetical to a low or high salience actor alters cognitive burden amongst our respondents. The second potential mechanism is differential treatment recall: that respondents are better able to recall treatments from salient actors than non-salient ones. Yet as Appendix §4 shows, we find no evidence that treatment recall rates significantly vary with the actor identity treatment. The third interpretation, which we believe is more consistent with our results, has to do with simple Bayesian models of persuasion: endorsement effects are stronger when the endorsement comes from a salient cuegiver because respondents are likely to have stronger priors about the cuegiver.¹⁹

4.4 CONTEXTUAL DETAIL EFFECTS

Lastly, we consider the moderating effects of contextual detail in Figure 4. We administered two versions of our context treatments. In the nuclear weapons experiment, respondents were either exposed to a reduced context vignette (baseline) or an original elaborate context vignette. In the in-group favoritism experiment, respondents were either exposed to an original minimal context vignette (baseline), or an extended context vignette which included “filler” or “charged” additional context. By focusing on the interaction effect of original treatments with contextual treatments, both panels of Figure 4 lend insight into the consequences of providing additional context in experimental vignettes.

As demonstrated in panel (a) in Figure 4, exposing respondents to the original rich experimental vignette in the Press, Sagan, and Valentino experiment has a negative moderating effect on the

¹⁹In this sense, our findings offer helpful scope conditions for other experimental work arguing that using unnamed or hypothetical actors artificially inflates the size of treatment effects (McDonald, 2019): if the dependent variable involves measuring attitudes about an actor, a simple Bayesian framework would predict that the stronger the respondents’ priors (as would be the case if the actor is salient), the *less* they should update in response to new information about the actor. However, if the dependent variable involves measuring attitudes about a policy, that same Bayesian framework would predict that the stronger the respondents’ priors about the policy’s endorser (as would be the case if the actor is salient), the *more* they should update in response to information about the cuegiver.

Figure 4: Adding contextual detail attenuates treatment effects

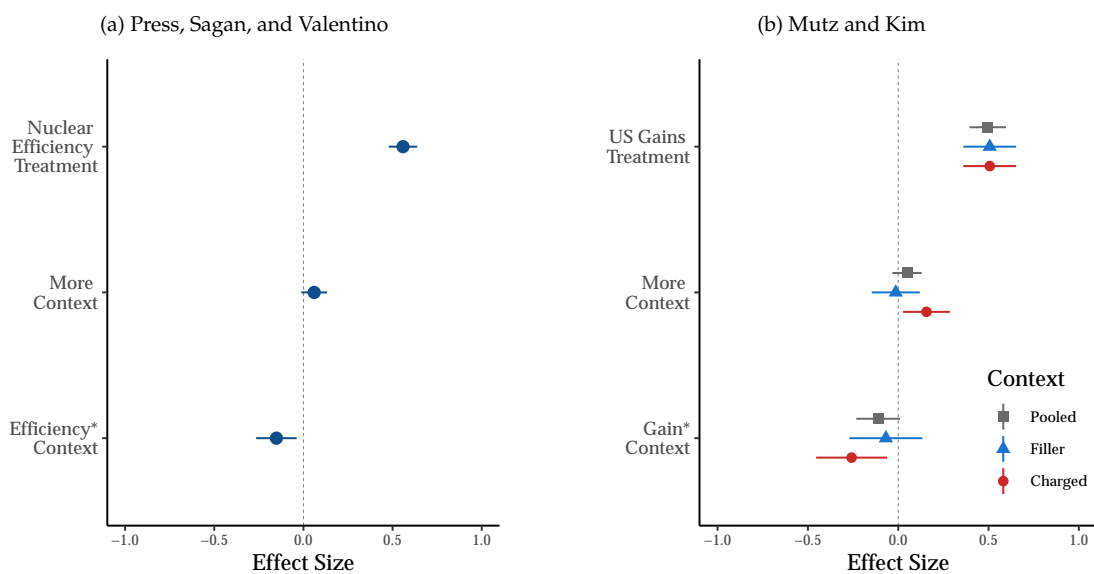


Figure 4 shows that adding contextual detail to studies weakens the treatment effects. Point estimates and corresponding confidence intervals are extracted from three separate OLS models where original outcomes are predicted by original treatments interacted with study level context. In panel *a*, a baseline reduced-context condition is compared with the original elaborate-context condition used by Press, Sagan, and Valentino. In panel *b*, we compare a baseline reduced context vignette with elaborate context conditions which are either non-innocuous (blue) or innocuous (red). We also consider a pooled model of both types of experimental context (gray). All outcomes are standardized.

study's main treatment. Put differently, extended experimental vignettes seem to dampen Press, Sagan, and Valentino's original treatment (nuclear effectiveness), but this moderating effect does not lead scholars to draw opposite inferences, but rather just estimate more conservative treatment effects.

Panel (b) in Figure 4 provides us with further insight in to the moderating effects of contextual detail on main treatments. In this panel, we consider the general effect of adding contextual detail to experimental vignettes (grey - pooled model), as well as the particular effects of adding either "filler" or "charged" context. Results from panel (b) further suggest that adding contextual detail to experimental vignettes will dampen treatment effects. Indeed, the moderating effect of extended contextual detail (in relation to a baseline minimal context condition), when pooling together both "filler" and "charged" context conditions, approaches statistical significance ($p < 0.08$). Clearly, as evident in panel (b) of Figure 4 this effect is driven by the "charged" context condition, which in and of itself has a statistically significant impact on the size (but not direction) of main treatment effects.

To better understand why adding contextual detail to experimental vignettes dampens original treatment effects, we consider the effects of our contextual detail treatment on treatment recall success. To do so, we regress respondents' recall success of the original study-level treatments (Nuclear attack effectiveness in Press, Sagan, and Valentino, and expected consequences of trade in Mutz and Kim) on respondents' contextual detail condition. Figure 5 demonstrates that increased context in experimental design hinders respondents' ability to successfully recall the treatment condition to which they were assigned. This provides insight as to the reason why increased context may dampen main treatment effects.

5 *Concluding Thoughts*

We began this paper by calling attention to a significant problem faced by political scientists who seek to test their theories using experimental designs: in most cases, they have a wide degree of latitude in how to design the experimental stimuli and must make choices about whether to use real actor names or make them up (or leave them un-named), whether to add rich, contextual detail (and if so, how much, and what kind), how to present the information in the experiment (whether explicitly hypothetical, implicitly hypothetical, or as real), whether to use deception, and so on. In

Figure 5: Contextual Detail Effects on Treatment Recall Success

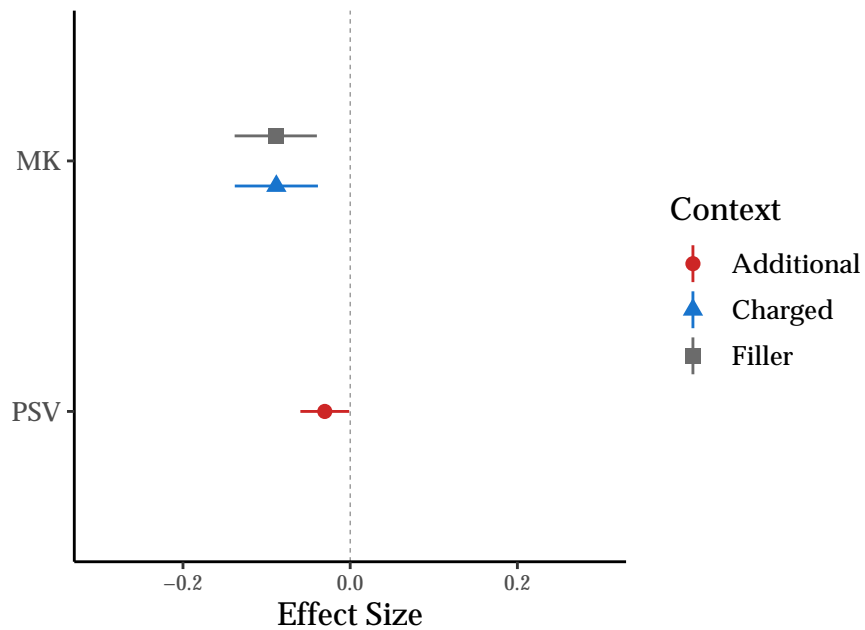


Figure 5 demonstrates how adding Contextual Detail Negatively Affects Treatment Recall. Point estimates and corresponding confidence intervals are extracted from three separate OLS models where a binary treatment recall success variable is predicted by the context condition to which respondents were assigned. The Press, Sagan and Valentino model (left) compares recall rates of respondents assigned to a baseline reduced context conditions, with respondents assigned to extended context condition. Mutz and Kim models, compare respondents assigned to a minimal baseline condition, with respondents assigned to innocuous (blue) and non-innocuous conditions. All outcomes are standardized.

confronting the issues raised by these “design degrees of freedom,” scholars have no shortage of strong folk wisdom to fall back on from their peers, mentors and textbooks, but the “conventional wisdom” on which they can rely is either nonexistent or contradictory. Specifically, the strong preference from those in the economics tradition is to use sparse, abstract designs, while those in the psychology tradition will prefer more rich context and detail and the use of real or made-up actor names. One thing nearly everyone agrees upon, however, is that — whatever our researcher chooses — they will face a trade-off between experimental control and generalizability. However, despite a recognition that these questions are, ultimately, subject to study and research like many other problems (e.g., [Friedman, Friedman and Sunder, 1994](#)), there is little in the way of theoretical frameworks or empirically-minded guidance for researchers who face these issues.

Our contribution here is twofold. First, we provided a conceptual framework that helps to make sense of the many choices that experimentalists face in terms of the degree of abstraction or concreteness of their designs. In particular, our framework outlines three dimensions of abstraction—situational hypothetically, actor identity and contextual detail—and helps us to classify prototypical experiments along these three dimensions. Most importantly, our framework and theoretical discussion of the implications of each of these three dimensions of abstraction for internal and external validity help to clarify a key point: the oft-remarked upon tradeoff between experimental control and generalizability is not nearly as stark as it is often assumed and in some cases is not a direct trade-off at all. Abstraction may in some cases enhance, rather than decrease, experimental control, which, in any case, experimentalists have less of than they realize in many cases.

Empirically, we test our theoretical framework through a replication and extension of three well-known vignette-based survey experiments in political science: [Nicholson’s \(2012\)](#) exploration of elite cues, [Press, Sagan and Valentino’s \(2013\)](#) test of the nuclear taboo, and [Mutz and Kim’s \(2017\)](#) study of the role of in-group favoritism in trade preferences. To each of these, we add our layers of experimental manipulations to test the implications of abstraction in experimental design. To [Nicholson \(2012\)](#), we manipulate the actor identity of politician presented in the vignette; to [Mutz and Kim’s \(2017\)](#) relatively sparse experiment we add two types of context (“filler” context and “charged” context likely to interact with the treatment) and to [Press, Sagan and Valentino \(2013\)](#) we add manipulations on the levels of both context and actor identity. In addition, for all three experiments, we manipulate the degree of situational hypothetically, presenting scenarios as either real, implicitly hypothetical, explicitly hypothetical, or without any mention of hypothetical-

ity.

Our empirical results suggest reasons for optimism. Framing a study as (implicitly or explicitly) hypothetical, or real does not make any substantial difference, failing to affect any of the main findings from the three replicated studies. This suggests that the difficult ethical decisions about whether or not to use deception may in many cases be unnecessary, adding empirical weight to an important normative debate in the field. We examined contextual detail in two ways: adding two types of context to [Mutz and Kim \(2017\)](#) and subtracting context from [Press, Sagan and Valentino \(2013\)](#) to render the scenario more abstract. Our results are consistent across both studies: in the vignette experiments we replicate here, we find that additional context leads to more conservative estimates of treatment effects, as long as the study has sufficient statistical power it is unlikely to affect the likelihood of finding an effect in the first place, and that context dampens treatments effects by hindering respondents' ability to successfully recall the main treatment. Choosing the appropriate level of contextual detail in experimental work thus depends on the purpose of the study: if the purpose of a study is to demonstrate that an effect exists, a sparser experimental design better enables researchers to identify the mechanism, but if the purpose is instead to understand how important an effect might be relative to other considerations, or whether respondents in a more naturalistic setting would be likely to receive the treatment ([Barabas and Jerit, 2010](#)), a more contextually-rich design may be beneficial.

We also investigated the effects of varying the level of abstraction of the actors in the experiments. We manipulated actor identity in the nuclear weapons experiment by exposing respondents to conditions in which the country was either unnamed (baseline), fictional ("Malaguay") or real and either consistent with the main thrust of the scenario ("Syria") or not ("Bolivia"), the consistency having been ascertained via an earlier pilot experiment. In the elite cues experiment, actor identity was manipulated using made-up, low-salience, or high-salience cue-givers. Across both experiments, which considered different types of actors (i.e. countries or politicians), most actor-related design choices did not matter, in that the interaction between the actor identity treatment and the main treatment was not statistically significant. The sole exception is that more salient politicians make more effective cuegivers than fictional cuegivers do. We also consider the extent to which different dimensions of our framework (contextual detail and actor identity) interact to moderate experimental findings, shown in Appendix §5. We find little support for this notion, further enhancing our intuition that decisions around actor identities do not substantively moderate

experimental findings.

In line with other recent work seeking to subject widely held assumptions about experimental methods to empirical scrutiny (Mullinix et al., 2015; Coppock, 2019; Mummolo and Peterson, 2019; Kertzer, 2020), we find limited empirical support to substantiate commonly held concerns regarding the consequences of design choices for the substantive interpretation of experiments in political science. Our conceptual framework clarifies where, when, and how researchers might have discretion in selecting particular levels of abstraction in their experimental stimuli. However, somewhat ironically, our evidence suggests that in cases where researchers have discretion over design choices relating to abstraction, their choices bear limited empirical consequences. Our null findings do not imply that levels of abstraction never moderate average treatment effects from experiments, but rather suggest that even if levels of abstraction moderate average treatment effects, they do so in a subtle manner which does not impact the substantive interpretation of any given experiment.

References

- Adida, Claire L. 2015. "Do African voters favor coethnics? Evidence from a survey experiment in Benin." *Journal of Experimental Political Science* 2(1):1–11.
- Aguinis, Herman and Kyle J Bradley. 2014. "Best practice recommendations for designing and implementing experimental vignette methodology studies." *Organizational Research Methods* 17(4):351–371.
- Alekseev, Aleksandr, Gary Charness and Uri Gneezy. 2017. "Experimental methods: When and why contextual instructions are important." *Journal of Economic Behavior & Organization* 134:48–59.
- Alexander, Cheryl S and Henry Jay Becker. 1978. "The use of vignettes in survey research." *Public opinion quarterly* 42(1):93–104.
- Arceneaux, Kevin. 2012. "Cognitive Biases and the Strength of Political Arguments." *American Journal of Political Science* 56(2):271–285.
- Banerjee, Abhijit, Donald P Green, Jeffery McManus and Rohini Pande. 2014. "Are poor voters indifferent to whether elected leaders are criminal or corrupt? A vignette experiment in rural India." *Political Communication* 31(3):391–407.
- Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2020. "Beyond the breaking point? Survey satisficing in conjoint experiments." *Political Science Research and Methods* Forthcoming:1–19.
- Barabas, Jason and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104(2):226–242.
- Berinsky, Adam J. 2009. *In time of war: Understanding American public opinion from World War II to Iraq*. Chicago, IL: University of Chicago Press.
- Boettcher, III, William A. 2004. "The prospects for prospect theory: An empirical evaluation of international relations applications of framing and loss aversion." *Political Psychology* 25(3):331–362.
- Boettcher III, William A and Michael D Cobb. 2006. "Echoes of Vietnam? Casualty framing and public perceptions of success and failure in Iraq." *Journal of Conflict Resolution* 50(6):831–854.
- Bracic, Ana and Amanda Murdie. 2019. "Human Rights Abused? Terrorist Labeling and Individual Reactions to Call to Action." *Political Research Quarterly* Forthcoming:1–15.
- Brader, Ted, Nicholas A. Valentino and Elizabeth Suhay. 2008. "What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration." *American Journal of Political Science* 52(4):959–978.
- Brooks, Deborah Jordan and Benjamin A Valentino. 2011. "A war of one's own: Understanding the gender gap in support for war." *Public Opinion Quarterly* 75(2):270–286.
- Burge, Camille, Julian J. Wamble and Rachel Cuomo. 2020. "A Certain Type of Descriptive Representative? Understanding How the Skin Tone and Gender of Candidates Influences Black Politics." *Journal of Politics* Forthcoming.
- 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.
- Butler, Daniel M, David W Nickerson et al. 2011. "Can learning constituency opinion affect how legislators vote? Results from a field experiment." *Quarterly Journal of Political Science* 6(1):55–83.
- Butler, Daniel M and Eleanor Neff Powell. 2014. "Understanding the party brand: Experimental evidence on the role of valence." *The Journal of Politics* 76(2):492–505.
- Camerer, Colin. 1997. Rules for experimenting in psychology and economics, and why they differ. In *Understanding Strategic Interaction*. Springer pp. 313–327.
- Cantor, Nancy and Walter Mischel. 1979. "Prototypes in Person Perception." *Advances in Experimental Social Psychology* 12:3–52.
- Clarke, Kevin A. and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of*

- Representations*. Oxford University Press.
- Colburn, Timothy and Gary Shute. 2007. "Abstraction in computer science." *Minds and Machines* 17(2):169–184.
- Colleau, Sophie M, Kevin Glynn, Steven Lybrand, Richard M Merelman, Paula Mohan and James E Wall. 1990. "Symbolic racism in candidate evaluation: An experiment." *Political Behavior* 12(4):385–402.
- Converse, Jean M and Stanley Presser. 1986. *Survey questions: Handcrafting the standardized questionnaire*. SAGE Publications.
- Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3):613–628.
- Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information equivalence in survey experiments." *Political Analysis* 26(4):399–416.
- Dawes, Christopher T, Peter John Loewen and James H Fowler. 2011. "Social preferences and political participation." *The Journal of Politics* 73(3):845–856.
- Dickson, Eric S. 2009. "Do Participants and Observers Assess Intentions Differently During Bargaining and Conflict?" *American Journal of Political Science* 53(4):910–930.
- Dickson, Eric S. 2011. Economics vs. Psychology Experiments: Stylization, Incentives, and Deception. In *Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge University Press.
- 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.
- Druckman, James N and Kjersten R Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47(4):729–745.
- Dunning, Thad and Lauren Harrison. 2010. "Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in Mali." *American Political Science Review* 104(1):21–39.
- Evers, Miles M, Aleksandr Fisher and Steven D Schaaf. 2019. "Is There a Trump Effect? An Experiment on Political Polarization and Audience Costs." *Perspectives on Politics* 17(2):433–452.
- Friedman, Sunder, Daniel Friedman and Shyam Sunder. 1994. *Experimental methods: A primer for economists*. Cambridge University Press.
- Gaines, Brian J, James H Kuklinski and Paul J Quirk. 2007. "The logic of the survey experiment reexamined." *Political Analysis* 15(1):1–20.
- 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.
- Hashtroudi, Shahin, Sharon A Mutter, Elizabeth A Cole and Susan K Green. 1984. "Schema-consistent and schema-inconsistent information: Processing demands." *Personality and Social Psychology Bulletin* 10(2):269–278.
- Herrmann, Richard K, Philip E Tetlock and Penny S Visser. 1999. "Mass public decisions on go to war: A cognitive-interactionist framework." *American Political Science Review* 93(3):553–573.
- Horowitz, Michael C and Matthew S Levendusky. 2011. "Drafting support for war: Conscription and mass support for warfare." *The Journal of Politics* 73(2):524–534.
- Huddy, Leonie and Nayda Terkildsen. 1993. "Gender stereotypes and the perception of male and female candidates." *American Journal of Political Science* pp. 119–147.
- Hundley, Lindsay. 2020. "The Shadow of the Future and Bargaining Delay: An Experimental Approach." *Journal of Politics* 82(1):378–383.
- Johns, Robert and Graeme AM Davies. 2012. "Democratic peace or clash of civilizations? Target states and support for war in Britain and the United States." *The Journal of Politics* 74(4):1038–1052.
- Kam, Cindy D and Elizabeth J Zechmeister. 2013. "Name recognition and candidate support."

- American Journal of Political Science* 57(4):971–986.
- 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. 2020. “Re-assessing Elite-Public Gaps in Political Behavior.” *American Journal of Political Science* Forthcoming.
- 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.
- Kreps, Sarah and Stephen Roblin. 2019. “Treatment format and external validity in international relations experiments.” *International Interactions* Forthcoming.
- Kriner, Douglas L and Francis X Shen. 2014. “Reassessing American casualty sensitivity: The mediating influence of inequality.” *Journal of Conflict Resolution* 58(7):1174–1201.
- Lau, Richard R, Lee Sigelman and Ivy Brown Rovner. 2007. “The effects of negative political campaigns: a meta-analytic reassessment.” *Journal of Politics* 69(4):1176–1209.
- LeVeck, Brad L., D. Alex Hughes, James H. Fowler, Emilie M. Hafner-Burton and David G. Victor. 2014. “The Role of Self-Interest in Elite Bargaining.” *Proceedings of the National Academy of Sciences* 111(52):18536–18541.
- LeVeck, Brad L. and Neil Narang. 2017. “The Democratic Peace and the Wisdom of Crowds.” *International Studies Quarterly* 61(4):867–880.
- Levine, David K and Thomas R Palfrey. 2007. “The paradox of voter participation? A laboratory study.” *American Political Science Review* 101(1):143–158.
- Lupton, Danielle L. 2019. “The External Validity of College Student Subject Pools in Experimental Research: A Cross-Sample Comparison of Treatment Effect Heterogeneity.” *Political Analysis* 27(1):90–97.
- 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.
- Mattes, Michaela and Jessica L. P. Weeks. 2019. “Hawks, Doves and Peace: An Experimental Approach.” *American Journal of Political Science* 63(1):53–66.
- 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. 2019. “Avoiding the Hypothetical: Why “Mirror Experiments” are an Essential Part of Survey Research.” *International Journal of Public Opinion Research* Forthcoming.
- Morton, Rebecca B and Kenneth C Williams. 2010. *Experimental political science and the study of causality: From nature to the lab*. New York, NY: Cambridge University Press.
- Mullinix, Kevin J, Thomas J Leeper, James N Druckman and Jeremy Freese. 2015. “The generalizability of survey experiments.” *Journal of Experimental Political Science* 2(2):109–138.
- Mummolo, Jonathan and Erik Peterson. 2019. “Demand effects in survey experiments: An empirical assessment.” *American Political Science Review* 113(2):517–529.
- 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.
- Nielson, Daniel L., Susan D. Hyde and Judith Kelley. 2019. “The elusive sources of legitimacy beliefs: Civil society views of international election observers.” *The Review of International Organizations* 14(4):685–715.

- Paivio, Allan. 1990. *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.
- 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.
- Quek, Kai. 2017. "Rationalist Experiments on War." *Political Science Research and Methods* 5(1):123–142.
- Raffler, Pia. 2019. "Does political oversight of the bureaucracy increase accountability? Field experimental evidence from an electoral autocracy." Working paper.
- Reeves, Andrew and Jon C. Rogowski. 2018. "The Public Cost of Unilateral Action." *American Journal of Political Science* 62(2):424–440.
- Reiley, David. 2015. The lab and the field: empirical and experimental economics, by David Reiley. In *Handbook of experimental economic methodology*, ed. Guillaume R Fréchette and Andrew Schotter. Oxford University Press, USA pp. 410–412.
- Renshon, Jonathan. 2015. "Losing Face and Sinking Costs: Experimental Evidence on the Judgment of Political and Military Leaders." *International Organization* 69(3):659–695.
- 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.
- Rosenwasser, Shirley M, Robyn R Rogers, Sheila Fling, Kayla Silvers-Pickens and John Butemeyer. 1987. "Attitudes toward women and men in politics: Perceived male and female candidate competencies and participant personality characteristics." *Political Psychology* pp. 191–200.
- Rousseau, David L and Rocio Garcia-Retamero. 2007. "Identity, power, and threat perception: A cross-national experimental study." *Journal of Conflict Resolution* 51(5):744–771.
- Rubenzler, Trevor and Steven B Redd. 2010. "Ethnic minority groups and US foreign policy: examining congressional decision making and economic sanctions." *International Studies Quarterly* 54(3):755–777.
- Sanbonmatsu, Kira. 2002. "Gender stereotypes and vote choice." *American Journal of Political Science* 46(1):20–34.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64(4):1033–1053.
- Semin, Gün R and Klaus Fiedler. 1988. "The cognitive functions of linguistic categories in describing persons: Social cognition and language." *Journal of Personality and Social Psychology* 54(4):558–568.
- Shapira, Oren, Nira Liberman, Yaacov Trope and SoYon Rim. 2012. Levels of mental construal. In *SAGE Handbook of Social Cognition*, ed. Susan T. Fiske and C Neil Macrae. SAGE Publications pp. 229–250.
- Steiner, Peter M, Christiane Atzmüller and Dan Su. 2016. "Designing valid and reliable vignette experiments for survey research: A case study on the fair gender income gap." *Journal of Methods and Measurement in the Social Sciences* 7(2):52–94.
- Teele, Dawn Langan, Joshua Kalla and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112(3):525–541.
- Tingley, Dustin. 2017. "Rising Power on the Mind." *International Organization* 71(S1):S165–S188.
- Tingley, Dustin and Barbara Walter. 2011a. "Reputation building in international relations: An experimental approach." *International Organization* 65(2):343–365.
- Tingley, Dustin H and Barbara F Walter. 2011b. "The effect of repeated play on reputation building: an experimental approach." *International Organization* 65(2):343–365.
- Tomz, Michael. 2007. "Domestic audience costs in international relations: An experimental approach." *International Organization* 61(4):821–840.
- Tomz, Michael, Jessica Weeks and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions about Military Force in Democracies." *International Organization* 74(1):119–143.

- Tomz, Michael R and Jessica LP Weeks. 2013. "Public opinion and the democratic peace." *American political science review* 107(4):849–865.
- Trager, Robert F and Lynn Vavreck. 2011. "The political costs of crisis bargaining: Presidential rhetoric and the role of party." *American Journal of Political Science* 55(3):526–545.
- Trope, Yaacov and Nira Liberman. 2003. "Temporal Construal." *Psychological Review* 110(3):403–421.
- Valentino, Nicholas A, Fabian G Neuner and L Matthew Vandenbroek. 2018. "The changing norms of racial political rhetoric and the end of racial priming." *The Journal of Politics* 80(3):757–771.
- Wamble, Julian J. 2020. "The Chosen One: How Community Commitment makes Certain Representatives More Preferable." Working paper.
- White, Ariel, Anton Strezhnev, Christopher Lucas, Dominika Kruszewska and Connor Huff. 2018. "Investigator characteristics and respondent behavior in online surveys." *Journal of Experimental Political Science* 5(1):56–67.
- Yarhi-Milo, Keren, Joshua D. Kertzer and Jonathan Renshon. 2018. "Tying Hands, Sinking Costs, and Leader Attributes." *Journal of Conflict Resolution* 62(10):2150–2179.