

How to make elite experiments work in International Relations

European Journal of
International Relations
2021, Vol. 27(2) 596–621
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1354066120987891
journals.sagepub.com/home/ejir



Simone Dietrich

University of Geneva, Switzerland

Heidi Hardt 

University of California, Irvine, USA

Haley J. Swedlund 

Radboud University, Netherlands

Abstract

For decades, many International Relations (IR) scholars did not engage in elite experiments, because they viewed it as too risky, too costly, or too difficult to implement. However, as part of a behavioral turn in IR, a growing number of scholars have begun to adopt the method in their own research. This shift raises important questions. Under what conditions do elite experiments add value to IR scholarship? How can scholars overcome the logistical and ethical challenges of sampling such an elusive group? This article makes an original conceptual contribution to methodological debates on the role of behavioral approaches by analyzing experiments on foreign policy elites. We analyze the method's strengths and weaknesses, evaluate ethical considerations, and present what is—to the best of our knowledge—the most comprehensive set of implementation guidelines. Our article draws on recently published IR research and argues that the payoffs from elite experiments are well worth the effort.

Keywords

Elite experiments, methodology, pluralism, foreign policy, experiments, International Relations

Corresponding author:

Haley J. Swedlund, Radboud University, Heyendaalseweg 141, Nijmegen, 6525 AJ, Netherlands.

Email: h.swedlund@fm.ru.nl

Introduction

Scholars in International Relations (IR) increasingly employ experiments on foreign policy elites despite the challenges associated with implementation (e.g. Mintz et al., 2006; Bethke (2016); Findley et al 2017a, 2017b; Hardt, 2018a, 2018b, Swedlund, 2017a; Hafner-Burton et al., 2017; Saunders, 2018; Yarhi-Milo et al., 2018; Busby et al., 2019; Renshon 2015; Tomz et al., 2020; Dietrich, 2021). Previous studies have criticized the method as too costly, too difficult, or simply too risky to carry out (Hafner-Burton et al., 2013; Peabody, 1990). Thus, most scholarship comprising IR's growing body of experimental research—that is, IR's “behavioral revolution” (Davis and McDermott, 2020; IO, 2017; ISR, 2007)¹—uses non-elite samples like students, the public, and Amazon Mechanical Turk (i.e. Redd, 2002; Kertzer, 2017; Findley et al., 2013; Bayram and Holmes, 2019). However, findings from elite experiments should be better able to approximate elite behavior; research suggests that non-elite experiments do not consistently predict elites' actual decisions and actions (e.g. Druckman and Kam, 2011). Given that elite experiments do have certain advantages, when is the method worth the costs?

In this article, we identify a set of conditions under which elite experiments add value to IR research, and we introduce what we believe to be the first set of practical guidelines for implementation. We offer IR scholars explicit guidance on navigating upfront investments, ethically accessing and recruiting a sufficiently large sample of elites, and deploying the experiment at minimum cost. Drawing on successful and ethically conducted examples, we posit that elite experiments elicit new and important findings due to the markedly different ways that elites, relative to non-elites, behave, including how they calculate risks, make decisions, and respond to incentives (Mintz et al., 2006; Tetlock, 2005).

We employ Hafner-Burton et al.'s (2013: 369) definition of elites as core decision-makers who (a) occupy “top positions in social and political structures,” (b) “have the highest indices in their branch of authority,” and (c) “exercise significant influence over social and political change” (also Pareto, 1935; Pakulski, 2008). Examples of elites include individuals occupying high-level positions in states and international institutions (e.g. members of legislatures, senior judges, ambassadors, military representatives in foreign and defense ministries, assistant secretaries general) and individuals operating at the highest levels (e.g. presidents, prime ministers and secretaries general).

Existing methodological scholarship has not yet specified when elite experiments can help scholars address their research questions. Related work has assessed the value and logistics of other types of experiments (e.g. cognitive neuroscience, natural experiments, etc.) (see typology in Mintz et al., 2011; McDermott, 2002a, 2002b, 2011; Gerber and Green, 2012). Additionally, several leading journals have devoted special issues to exploring experimental research in IR (IO, 2017; ISQ, 2011). However, this body of work has not investigated elite experiments as a unique type of experimental research.

Therefore, this article makes a conceptual contribution to existing scholarship on methodological pluralism in IR by providing a comprehensive presentation of practical advice for employing elite experiments as a methodology. We strongly support methodological openness and encourage the use of elite experiments as complementary to traditional methods. With respect to scope, we focus mainly on interview and survey-based experiments, excluding field experiments (e.g. those involving randomized evaluations

in international institutions). While the latter are important for the field of IR, their purpose differs (Hyde, 2010; 2015; Loewen et al., 2010).

The article proceeds as follows. First, we articulate why surveying elites is uniquely valuable for IR scholars. Second, we specify the conditions under which elite experiments are likely to be most useful and not useful. In this section, we also discuss what elite experiments can provide that other methods cannot but also how elite experiments complement other methods. Third, we provide guidelines for IR scholars new to the method and to those seeking to improve implementation of the method.

A rationale for conducting experiments on elites

Historically, theories of IR have focused on the behavior of states (and more recently non-state actors)—often neglecting the role of decision-makers themselves. Over the decades, however, many IR scholars have increasingly moved away from classical paradigm debates (e.g. realism, liberalism) (Dunne et al., 2013), arguing that the field should focus on identifying and testing casual mechanisms and developing mid-range theory (Bennett, 2013). Coming out of these debates, more and more IR scholars are focused on understanding of how individual-level preferences and beliefs affect decision-making (Hafner-Burton et al., 2017). In this regard, survey and interview-based experiments with foreign policy elites offer a promising tool. In addition to permitting strong causal inferences, experiments with elites can help us to better understand the effects of individual-level heterogeneity on choice processes and strategic interactions. In the words of one IR scholar, “[F]or a subset of research agendas that focus in part on elite decision making, conducting experimental research directly with elites is one way to make experiments in IR more realistic and potentially more relevant” (Hyde, 2015: 409).

Sampling foreign policy elites makes intuitive sense since, in IR research, these elites are often the scholar’s population of interest. However, scholars have cautioned against elite experiments because of perceived and, sometimes, actual barriers to access and costs associated with implementation. Hafner-Burton et al. (2013: 368) argue that it is hard to conduct experiments on experienced elites, “because they are generally busy, wary of clinical poking, and skittish about revealing information about their decision-making processes and particular choices.” Additionally, Peabody (1990: 452) argue that because elites are farther up than local or regional elites in a government’s chain of command, accessing them may be even “more difficult” in the subfield of IR than other political science fields. Given the potential risks and costs, why not exclusively sample non-elites (e.g. college students, MTurk respondents, etc.)?

We believe that scholars should consider investing in elite experiments precisely because research suggests that elites behave fundamentally differently from non-elites. Consequently, elite experiments can and should be able to provide us new, different, and interesting results relative to those conducted on non-elite samples. Scholars in other sub-fields of political science have successfully incorporated elite experiments as a methods,² and the value of elite experiments is also increasingly acknowledged in other fields like Public Administration where scholars have turned to examine experimentally the individual attitudes and behavior of bureaucrats on issues as varied as accountability or

professional behavior or performance (Baekgaard and Serritzlew, 2016; Grohs et al., 2016; Jilke et al., 2016). We thus situate our article in a broader movement that recognizes the advantages of conducting elite experiments.

Elites as agents of disproportionate influence

There are several reasons to believe that the behavior of foreign policy elites is different from the general public. First, as agents of foreign policy, elites have an advantage over non-elites in their ability to make decisions. Their extensive practical experience makes them efficient decision-makers. Unlike non-elites, experience aides them in selecting appropriate heuristics that are tailored to their specific circumstances and improves their credibility in signaling (Hafner-Burton et al., 2013: 374).

Second, because they are “less prone to loss aversion,” elites appear to be better at taking risks. Years of experience as a politician—whether as an ambassador or prime minister—hone one’s political skills (Hafner-Burton et al., 2013). Not all elites, however, behave in the same way. The ability to use bargaining skills to successfully predict future outcomes is tempered by the type of elite one is. Building on Berlin’s distinction between “hedgehogs” and “foxes,” Tetlock (2005), for example, argues that elites who are experts in many areas and draw from many sources of knowledge are more likely to be able to correctly predict events than those who are more specialized.

Third, relative to non-elites, elites are confronted with significantly more information inputs. According to Axelrod (2015), this can lead elites to consider “more relevant beliefs than they can handle,” limiting cognitive capacity and resulting in elites “simplifying” how they perceive policy decisions. Relative to non-elites, elites may in fact be no better at making predictions than non-experts (Tetlock, 2005). However, the consequences of their overconfidence can be deadly (Blainey, 1988; Jervis, 1976).

Fourth, beyond their different ways of approaching and taking political decisions, elites also have agent-centric differences in their attitudes and interactions with different types of groups. Previously, scholars found that, on most issue-areas, elites tend to be more tolerant than non-elites because of their experience with socialization (Stouffer, 1956; Sullivan et al., 1993). The election of more conservative and far-right governments in the US and Europe suggests that IR scholars may need to seek new methods for assessing elites’ tolerance. Nevertheless, elites’ political attitudes inform scholars about elite behavior, because attitudes signal how the elites interpret actions. These interpretations indicate how elites “guide the definition of problems” in their respective organizations (Aberbach et al., 1975).

Unique structural constraints on elite behavior

In addition to inherent differences in elite behavior, foreign policy elites are also likely to face different structural constraints than the general public. Within states, governing elites must maintain the support of and satisfy the preferences of key groups that are critical for those elites to stay in power (Mesquita et al., 2003). The type of political regime in which an elite is situated can alter his or her behavior. In democracies, political elites’ careers are dependent on reelection, which constrains their ability to act independently. The field of American politics has extensively examined how responsive political

elites are to constituent communication (Costa, 2017). Yet, just as democracies constrain elite behavior (Rohrschneider, 1994), autocracies do too (Weeks, 2014).

Elites also face structural constraints in international institutions. In international organizations (IOs), ambassadors and military representatives have a limited bargaining range, which is constrained by instructions provided by state capitals (Abbott et al., 2015; Hardt, 2014). Even though states delegate certain functions to IO secretariat staff, elites such as secretaries-general and assistant secretaries-general remain beholden to the shifting preferences of the IO's member states (Hardt, 2016; Hawkins et al., 2006; Hooghe and Marks, 2014). Authority may exist but is nevertheless limited by the design of the IO itself. In nongovernmental organizations (NGOs), elites such as the heads of the International Red Cross or Doctors without Borders must pursue their mandates under the constraints of the preferences of donors, since they hold the purse strings (Milner, 2006).

Finally, norms can powerfully influence elite behavior. Changes in norms can affect elite attitudes toward human rights, strategies in war, the nature of war, and international law compliance (Fortna, 2015; Sandholtz and Stiles, 2009; Simmons, 2009). Relative to citizens, state elites (e.g. presidents, prime ministers, foreign ministers, ambassadors) may also be relatively more responsive to international norms due to their increased interaction with other elites through IOs (Finnemore, 1993; Risse and Sikink, 1999; Simmons, 2009: footnote 94 on 138; Checkel, 2001: 2005). Busby et al. (2019), for example, find that US foreign policy elites are more responsive to multilateral approval than the US public. Additionally, institutional biases against certain groups (e.g. women, ethnic minorities, religious minorities), which are embedded in laws and norms, can affect how elites lead (Caprioli and Boyer, 2001; Jalalzai, 2008; Reynolds, 1999). For example, in Afghanistan, gendered norms of men as belligerents and women as pacifists have limited US foreign policy elites' ability to prevent terrorist attacks by female ISIS recruits (Bloom, 2012).

The methodological value of elite experiments

What exactly are elite experiments, and when and where do we expect the method to be most useful to IR scholars? We define experimental research broadly as research where scholars retain control over random assignment of experimental conditions in order to uncover causal relationships. When implemented properly, no other methodology is able to offer the robust support for causal claims that experiments do. McDermott (2002a) lists standardization and randomization as important experimental design aspects that all experimentalists need to consider including. Applied to elite experiments, standardization requires that researchers administer the same experimental protocol in the same way to their elite subjects across experimental conditions. Randomization refers to the random assignment of elites to the various experimental conditions to ensure that no unrelated or spurious factor biases the results.

Experimental design can be either between-subjects or within subject (McDermott, 2002a). In between-subject designs, elites are randomly assigned to experimental and control conditions and researchers compare experimental conditions to one another or a control condition. In within-subject designs, elites are randomly assigned across all experimental

or control conditions, thus serving as their own control and enabling experimenters to measure the treatment effect by comparing elite responses across the conditions.

In some instances, the treatment will dictate the choice of design, but often in elite experiments both designs are possible (McDermott, 2002a). One advantage of between-subject designs is that elite subjects, who are only exposed to one experimental condition, do not engage in learning or knowledge transfer across the conditions. Another advantage is that random assignment to one experimental condition only makes for a relatively short session compared to within-in subject design experiments.

Within-subject designs, on the other hand, are advantageous because they require fewer participants than between-subject designs (Dietrich, 2021). Each member of the elite responds to all experimental (including control) conditions and is thus part of all treatment and control samples. If a researcher has a total of four experimental conditions, for example, the sample counts four data points per individual elite. For a between-subject design, the research would require four times as many elites to get the same number of data points. Within-subject design also minimizes random noise. Despite random assignment into different treatment conditions, differences among elites may persist, affecting their responses and making it more difficult to detect significant treatment effects. Potential factors that can contribute to these individual differences even after randomization include, for example, elites' own history, their background knowledge, and the survey-taking context. In our view, the choice of design requires the consideration of multiple factors including theory of decision-making as well as practical considerations like the accessibility and approachability of elites—an issue that we will return to in the next section.

Elite experiments should add the most value to studies aimed at explaining the behavior of the elite decision-makers themselves in state and international institutions. For example, questions that address racial or gender biases in IOs are often difficult to address with conventional methods because they require elites to be forthcoming and aware of these biases. List experiments and scenario-based experiments with elites can allow scholars to indirectly assess such potential biases (Swedlund, 2017a). An elite experiment could help address questions such as whether or not bureaucrats are truly objective or whether they are supporting the national interests of their countries of origins. Such experiments can also be useful for examining conditions under which elites practice compliance (e.g. with international law or international norms). Scenario-based experiments with elites could also help testing competing theories for when institutions do and do not cooperate with one another. Additionally, elite experiments can also be run on diplomats to determine conditions under which international negotiations are more or less likely to break down.

By contrast, we can also consider the types of studies where elite experiments likely would *not* add much value. These would include research aimed at explaining any or all individuals' political behavior. In these cases, sampling from the general population would be sufficient. Similarly, elite experiments may not be particularly useful for studies assessing directly observable state behavior (e.g. voting in UN General Assembly, economic sanctions, conflict initiation). In the latter situations, scholars could collect data from voting records, government websites and archives, and existing conflict and war databases. Finally, we recognize that certain elites may typically be

too inaccessible (e.g. presidents, prime ministers), although access often varies dramatically depending on the country in question. In most instances, other types of elites (e.g. presidential advisors, diplomats, cabinet members, high-level military officials) can serve as useful proxies. Below, we discuss five ways in which elite experiments add value to the field of IR.

Facilitating theory testing on an essential sample

First, elite experiments allow scholars to test hypotheses on a sample that is often an essential sample for questions regarding state and IO behavior. Research on foreign policy has often focused on public perceptions and thus the *impact* of foreign policy, rather than the actual *decisions* of foreign policy elites (Sasley, 2010: 687). Although often efficient, lab experiments with student samples or convenience sampling (e.g. via Amazon MTurk) can challenge inferences in at least three ways: (1) the causal effect of the treatment, (2) the effect size, and (3) the heterogeneity of effects (Druckman and Kam, 2011). Levitt and List (2007) demonstrate that lab subjects' knowledge of being evaluated in the lab, coupled with the self-selection of volunteers for experiments limit our ability to infer to a wider population. Furthermore, as shown by Hafner-Burton et al. (2013) and Fehr and List (2004), there are significant differences between undergraduate subjects and more realistic subject populations.

Particularly worrying is the practice of "perspective-taking," which asks ordinary citizens to "act as if" they assumed a leading role in decision-making. Rather than mirroring decision-making by elites, citizen subjects have been found to act in ways that represent an exacerbation of pre-existing attitudes (Kertzer and Renshon, 2015). In one study, Kertzer and Renshon found that subjects hawkish toward the use of force became more hawkish, while their dovish counterparts became more dovish. In other words, in the lab, "perspective-taking thus makes people act more like themselves", rather than mirroring decision-makers (2015).

When asked to engage in perspective taking, subjects confront decision-making situations for the very first time with only limited knowledge and response time at their disposal. For example, Bayram and Graham (2017) employed a convenience sample of students to test hypotheses on the determinants of governmental funding choices in IOs. Student subjects adopted the perspective of a "top-level foreign policy advisor to the President," and then had to make funding recommendations based on hypothetical information.³ The design of the treatments is oversimplified,⁴ presumably to enable decision-making for student subjects in the first place. Consequentially, the results are not surprising. Nevertheless and despite a disclaimer about external validity, the study makes broad inferences about foreign policy elites' behavior.⁵

Strategic decision-making on global issues requires elites to have sufficient knowledge and experience of not only the issue area but also the decision-making process. Students and/or members of the general public are simply not likely to have such knowledge. A recent study involving an elite experiment on current and former members of Israel's legislature illustrates this advantage. When seeking to study decision-makers' beliefs about resolve in international crises, Yarhi-Milo et al. (2018: 2159) argue,

the most direct way to do so is to sample from exactly that population: leaders who have had to wrestle with these issues outside the lab. This is, after all, what is unique about studying leaders rather than the general public. . . our elite participants have repeatedly encountered these issues. . .

While there are many advantages of a drawing on an elite population, elite samples certainly do not guarantee external validity. In another study with Israeli parliamentarians, Tomz et al. (2020) investigate the role of public opinion in shaping foreign policy decisions, capitalizing on experiments with both Israeli parliamentarians and Israeli and US citizens. The study provides empirical data on an important but controversial question in foreign policy: does public opinion matter for foreign policy? While incredibly informative, the study's findings may or may not transfer to political contexts beyond Israel or the US. To ensure generalizability, the study would need to be replicated in other contexts.⁶

Expanding access to evidence by overcoming social desirability bias

Second, elite experiments can allow scholars to access evidence about elites' true biases and preferences that would be otherwise difficult to reveal. Elites, typically acting in an official capacity, are frequently under significant pressure to represent themselves, their state, and/or their institution in a positive light. Therefore, when asked directly about sensitive topics using non-experimental measures (e.g. surveys, questionnaires, interviews, focal groups), elites are likely to be hesitant to reveal pertinent information that, while accurate, would risk their reputations, let alone their careers. As Aberbach and Rockman (2002: 674) observed, elites "do not like being put in the straightjacket of close-ended questions." Questions on politically sensitive subjects (e.g. sexual violence in the military, government responses to environmental disasters, preferences in trade negotiations, gender bias in peace negotiations) are to be answered with "no comment," if a scholar is lucky enough to securing a meeting with the elite in the first place.

Experimental approaches like list experiments or scenario questions allow researchers to get at these topics in less confrontational ways and can be fruitful in circumventing social desirability pressures (Lavrakas, 2014). For example, Swedlund (2017a) sought to explain when foreign aid agencies were likely to suspend aid for political reasons. Prior research had only been able to study this issue indirectly by examining whether or not annual foreign aid disbursements increase or decrease in response to general trends in national-level governance measured by indexes such as Freedom House or Polity. However, this approach is not able to capture how foreign aid agencies respond to specific political transgressions.

An original survey of high-level donor officials working across twenty countries in sub-Saharan Africa presented her with the opportunity to directly query elites about the likelihood of aid suspensions (Swedlund, 2017b). Still, she was concerned that respondents would not voluntarily provide information that painted their agency in a bad light. Therefore, in order to elicit truthful responses about the conditions under which donors are willing to suspend foreign aid, Swedlund devised a list experiment, which she embedded into her survey. Combined with additional survey data, the list experiment allowed to test key hypotheses on the willingness of donor agencies to enforce political conditionality (Swedlund, 2017a). Testing these hypotheses would not have been possible

without directly querying elites, and the experimental approach helped limit concerns about social desirability bias.

Another example of elite experiments overcoming social desirability bias is Bayram's (2017) study of the relationship between German parliamentarians' cosmopolitan identities and the extent to which they feel obligated to comply with international law. If the scholar were to interview parliamentarians one-on-one or as part of a focal group, they would likely feel inclined to state that they of course support compliance with international law. Respondents would likely want to protect the reputation of their government, resulting in limited variation. Therefore, Bayram provided parliamentarians with surveys in which scenarios varied by conditions (as embedded treatments). Asking parliamentarians to respond to hypothetical scenarios allowed for testing the impact of cosmopolitan social identity on compliance—without putting elites in a situation in which they felt obligated to provide a certain type of response.

Encouraging the specifying of theories

Third, elite experiments can compel scholars to fully specify the theories that they seek to test in at least two ways. On the one hand, elite experiments encourage scholars to understand and test the individual-level implications for their argument. On the other hand, they allow scholars to test theories that consider the public to be an important driver of foreign policy and IR. In an elite experiment in foreign policy, Mintz et al. (1997) sought to specify existing theories about the decision strategies that elites use in times of crisis. In a computer-based experiment involving 44 top-ranking US Air Force officers, the authors presented the subjects with a dynamic choice set and were ultimately able to theorize and specify which types of strategies are used in a decision process at which times and the order in which they were used. The authors found that alternative-based expected utility strategies tend to operate more in the second stage of a decision process.

As another example, Dietrich (2016; 2021) studied the effect of donor political economies on donor decisions about how to deliver foreign aid abroad. The study sought to understand why, given that recipient countries' institutions experience corruption and inefficiencies, "some foreign aid donors outsource the delivery of aid to non-state actors" whereas "other donors continue to support state management of aid" (2016: 65). Dietrich theorized that differences in national orientations concerning what the state's role should be in public service delivery could explain variation in donor decisions. In addition to using dyadic time-series cross-section aid allocation data, the author employed a within-subject elite survey experiment among elites from six donor states to test the theory's individual-level implications. Results indicated that decision-makers across the two types of political economies proposed significantly different time horizons with which they assessed project success within their agencies.

In the aforementioned study, empirical support for individual-level implications corroborated the plausibility of the theory since it specified expectations across different levels of analysis. Once state theories are specified at the level of the individual decision-maker, scholars can derive additional empirical implications from the results of such

elite experiments; these implications may reveal interesting variation in how preferences are aggregated across agencies to explain state-level or system-level outcomes.

Relatedly, elite experiments can inform and help us test theories that consider the public to be an important driver of foreign policy and IR. In studies of foreign economic policy, for instance, scholars often assume that the “median voter” plays an important role in determining national preferences, or that stylized economic interests determine policy, as they may be filtered through a simple representation of legislative-executive relations. At the same time, ample evidence exists suggesting that the public does not have (strong) preferences and/or are not well informed on global issues (Eichenberg, 2016).

The latter raises several questions. Do voters punish incumbents at the ballot box for policy on global issues? Do elites believe that they will be punished by the electorate, if they do not take the public’s preferences seriously? Research with elites is essential to help us answer these types of important questions. Currently, we are aware of only one study that examines elite perceptions towards the role of the public in foreign policy (Tomz et al., 2020). Experimental studies of decision-makers can help improve our understanding of the conditions under which public preferences matter to decision-makers across global issue areas, allowing us to test theories about the role of public opinion in foreign policy decision making.

New hypotheses from experimental results

Fourth, elite experiments can contribute to the development of new hypotheses that would not be plausible if the focus was only on institutions, rather than elite preferences and behavior. As an example, one study sought to identify which sources influenced NATO elites (i.e. the population of interest) to be more or less likely to pass on knowledge of past strategic errors to their supervisors, colleagues, and successors (Hardt, 2018a). Hardt theorized about two sources of knowledge (i.e. treatments), the NATO secretariat and the US government, and conducted one-on-one meetings with 120 NATO civilian (e.g. ambassadors, secretariat staff) and military (e.g. high-level representatives) elites over 6 months. In each meeting, the elite completed a 5-min-long paper-and-pen survey experiment and then answered interview questions for the remaining time. Results indicated that, despite the US’ dominant role in NATO, knowledge received from the US deterred the elites from passing the knowledge down across time and space. Such findings prompted new hypotheses about why one would see a consistent negative effect across multiple models. In the work, Hardt suggests that elites were reticent because the US had politicized intelligence in the past (e.g. vis-a-vis weapons of mass destruction in Iraq). The author was then able to test new hypotheses on the coded responses to the interview questions that followed the experiment (Hardt, 2018a, 2018b).

In another example, Sheffer et al. (2018) were also able to derive new hypotheses that would otherwise be unlikely to be uncovered without having experimented on elites. The study involved conducting a series of experimental decision tasks and interviews on two samples: incumbent legislators from three democratic states and citizens from the same states. The study found surprisingly that—while politicians do make decisions differently from nonpoliticians, they are not always good at it. In fact, they may be even more

subject to making known choice anomalies as nonpoliticians. Findings like this encourage new hypotheses about elite decision-making, which can be tested in future research. For example, the author hypothesizes that a politician's conduct in office "may be the result of selection effects that increase the likelihood of specific types of people winning office and subsequently exhibiting more (or less) of a certain bias or heuristic in their decision making" (Sheffer et al., 2018: 317–318).

Here it is important to note that one drawback of elite experiments is that smaller sample sizes often prevent the inclusion of a number of cross-cutting treatments to test additional hypotheses. At the same time, promises to guarantee anonymity—which are often necessary to garner participation—may prevent researchers from testing additional hypotheses. For example, in the aforementioned study on aid suspensions, Swedlund (2017a) was unable to test for variation between different donor agencies because sample sizes across the different agencies were too small. Moreover, in order to convince respondents to complete the survey, she had to promise that she would not identify their institution in her research.

Bridging the gap between theory and practice

In addition to methodological benefits, elite experiments can also bridge the communications gap between academics and policymakers. More than two decades ago, George's (1993) "Bridging the Gap: Theory and Practice in Foreign Policy" called for the need to bring together academia and the policy world, while highlighting the methodological, political, and psychological obstacles in doing so. In 2002, Putnam, while cognizant of existing challenges, bemoans the academy's focus on theory and modeling and warns that political scientists risked marginalization, if they did not become more engaged with policymakers. Data from the 2011 and 2014 Teaching, Research and Policy (TRIP) survey of IR scholars in the US and in 19 other countries show that academics have become more engaged with members of policy circles. Thirty seven percent of scholars would like to be more engaged, with the demand to foster linkages being highest among scholars who study foreign aid and development, international political economy, and international security (Schneider, 2015).

We believe that testing hypotheses about state behavior on a sample of elites, who are relevant to the research question at hand, can contribute to bridging the gap between the academy and policy in two ways. First, it provides a platform for communication and exchange between the two camps, which, ideally, would create trust and possibly open up opportunities for substantive collaboration in research and consulting. Elite experiments force researchers to communicate their research plans and needs to policymakers, opening up a line of communication. Second, if the paper is about elites making foreign policy decisions, an empirical test of the theory on decision-makers makes the insights of the paper more relevant to the policy community. If elites do not find the questions and/or scenario in the experiment relevant to their day-to-day lives as decision-makers, the experiment will likely fail because elites may either decline to participate halfway through the experiment or may provide responses that lack meaning or not relevant to one's research question.

Practical advice for ethically conducting elite experiments

In this section, we provide scholars with practical guidelines on how to implement elite experiments while respecting and protecting human subjects. Scheduling meetings and successfully convincing elites to participate in surveys and/or answer verbal questions requires significant time, resources, and patience (Aberbach and Rockman, 2002; Loewen et al., 2010; Tansey, 2007). However, we believe that most barriers to accessing and carrying out experiments with elites are surmountable, if scholars employ appropriate techniques.

Investing time and resources to ensure sufficient sample size

For most elite experiments, and unlike convenience samples, investing significant time and resources into ensuring a sufficient sample size will be essential. A sufficient sample size is necessary for properly evaluating the treatment effects of the experiment. Consequently, for cases where travel is needed (e.g. if the experiment is embedded within in-person interviews), scholars should apply for funding well in advance. Similarly, scholars will need to set aside sufficient time in the field, at least a few months, in order to provide adequate time to ensure a medium-sized N . This advantages scholars with more resources, including more research time. Given the COVID-19 pandemic, experiments embedded as part of interviews are more likely to be needed to be conducted remotely. The authors have found elites, including US government officials, to be amenable to such accommodations.

Few elites in international institutions have contact information available online. Therefore, it may be necessary to start with whatever contact information is available and then send many emails and phone calls and in person requests in order to obtain a composite picture of the relevant sample. Moreover, it is inevitable that securing a sufficient sample size will require significant outreach, including sending multiple emails after a few days to non-respondents and following up with phone calls. (Depending on the type of institution [e.g. government], scholars should also consider sending their requests on letterhead by fax, in addition to email.) Overall, we have found that subject recruitment is likely to be most time-intensive part of the research process.

One option to increase recipient buy-in is to work with a well-established member of the policy community. In order to access a cross-section of US foreign policy elites for their experiment on multilateralism and the use of force, Busby et al. (2019) worked the Chicago Council on Global Affairs-Texas National Security Network. Working with members of the policy community may help increase response rates by lending greater policy credibility to the survey. However, collaboration with a policy institution can also present challenges, as policymakers will almost certainly have their own ideas about the content. This could make it more difficult to design the survey exactly as the researchers would like.

Offering financial and/or informational incentives at the time of subject recruitment may also help scholars achieve a sufficiently large sample size. In accordance with IRBs, scholars may wish to consider providing elites with Amazon gift cards, for example, to help boost response rates in the case of online or by mail survey experiments. However,

many elites, particularly administrative elites, are not allowed to accept payments in accordance with anti-corruption rules or laws. Alternatively (or in combination with a financial incentive), scholars can provide informational incentives to encourage participation in the experiment. In our research, we found that elites were eager to hear back about the results of our respective experiments. Often, elites explicitly asked us about when results would be ready even before the study began and indicated frustration about failed past attempts of trying to find out about study results from researchers. The main motivation for elite participation in research is often a belief that such research will help improve the functioning of their organization. Therefore, we encourage researchers to tap into this to encourage participation.

One option for scholars is to provide elites with policy briefings (either electronically, by phone and/or in person) that summarize the results of the study and policy recommendations. As we discuss in more detail below, scholars can also offer to provide presentations and/or just be a point of contact for policymakers on a given topic. Of course, scholars can simply send elites copies of their articles and/or books once the experiment is published. However, elites have preferences when it comes to the types of academic materials they are likely to read (e.g. Avey and Desch, 2014). Moreover, paywalls often prevent practitioners from accessing scholarly research. Therefore, full copies should be distributed (in line with copyright laws). Finally, scholars should be transparent with elites about how long it may take for a given study to be published.

Pre-specifying treatments in experimental design and ensuring a representative sample

Depending on their design, elite interviews allow scholars to take a more inductive approach. In contrast, elite experiments should be explicitly designed to test a pre-specified theory. Their value, like experiments more broadly, is precision in the treatment condition, allowing for strong causal inference. Therefore, scholars need to spend sufficient time considering which treatments to include and which will be excluded. Conducting the experiment ahead of time on a pilot sample is extremely beneficial. As already mentioned, relative to experiments embedded in public opinion surveys, elite experiments will have a much smaller *N*. The researcher will thus be limited in their ability to apply cross-cutting treatments during the analysis phase. This makes it all the more important that the necessary treatments conditions are pre-specified from the outset and power analyses are conducted to ensure that the sample size is large enough to test the treatment conditions the researcher is most interested in.

For elite experiments, the composition of an elite sample is as important as the sample size. Scholars should aim to create a representative sample by researching the demographics of the population (e.g. institution where the elites work) at length before beginning subject recruitment. In the cases of legislatures, the UN General Assembly or other bodies in which elites are elected or appointed officials, scholars can often easily determine the size of the population and which elites comprise it. However, an elite population may not always be known and/or information about the population (e.g. a phone book, registry, office directory, etc.) may not be publicly available. In these cases,

scholars can contact the institution's human resources office and request information about the population. Here we have found it extremely beneficial to already have contacts within a given population before undertaking a survey.

Having a breakdown of the population's characteristics (e.g. gender, age, nationality, years of experience, etc.) is important because certain covariates may affect experimental results. Scholars should consider which characteristics matter and research these ahead of time so that they can work to ensure that their sample approximates the population as much as possible. If a scholar knows, even in general terms, what the population's characteristics are then the scholar can apply a sampling strategy that targets specific members of the population.

Ethical considerations and protecting elites as human subjects

If available, we strongly recommend that scholars register their study with their academic institution's IRB or other relevant national ethics board before beginning subject recruitment.⁷ As the American Political Science Association Guide on Professional Ethics notes, "under certain conditions, political scientists are also legally required to assess the risks to human subjects" (APSA, 2012: 28). Unlike other types of human subject research in political science (e.g. interviews, surveys), experiments sometimes involve strategically omitting information and/or deception, which has prompted concerns regarding the impact on participants (e.g. voters receiving mailers prior to an election) (Basken, 2015). Even participating in a study without deception can sometimes put an elite in a vulnerable situation in terms of his or her employment or safety depending on the political and economic context in that country.

There are also a number of other ethical questions scholars ought to consider before deciding to proceed with an experiment on elites: How much time will the experiment take away from the elite's primary responsibilities? How, if any, will this affect a politician's responsibilities to meeting the needs of his constituents? Will the elite or his/her constituents benefit in any way from learning more about the topic of the study? Anecdotes circulate about local legislator fatigue from scholars sending list experiments that took time away from the needs of their office. Research also should be conscientious about lingering suspicions about experiments following several highly publicized cases of political science experiments that were condemned and/or retracted for unethical practices, such as the Montana voters' study (Bartlett, 2014) or the study on canvassers and voters' views on same-sex marriage (McNutt, 2015).

If the experiment is being conducted orally, scholars might consider the option of providing verbal informed consent rather than requiring written, signed consent forms.⁸ The former involves reading the elite a brief script that summarizes the research, tells him or her how the data will be used, and reminds him or her that participation is voluntary. If written consent is required, it is important to keep the consent form free of academic jargon so that the elite understands to what he or she is agreeing to.

There also exist multiple outlets for IR scholars to preregister experiments online. Such websites allow scholars to upload an analysis plan, including the hypotheses, treatments, controls, and other research design information, before conducting the experiment. As part of broader discussions on transparency in the social sciences, there exists an active scholarly debate on the utility of preregistration (Coffman and Niederle, 2015;

Kern and Gleditsch, 2017; Lin and Green, 2016). One of the costs of such a plan is that it may limit the scholar's ability to garner and implement new ideas for new hypotheses in the course of fieldwork (Olken, 2015: 62). However, Lin and Green (2016) argue that such plans can limit scholars from "tilting" a study "toward a desired result", lending more creditability to the findings.

Researching elites and institutions before beginning subject recruitment

When designing their study, scholars should spend a lot of time reading existing formal documentation, the website of the relevant institution(s), and relevant bodies of international law. Scholars should be as informed as possible about the elite, the elite's position in the institution, the structural constraints that the elite is likely to experience. They should also understand the role of different agencies within the international institutions. This requires that the researcher looks beyond the scholarly literature. In carrying out an elite experiment, one needs to avoid scholarly, technical jargon. Researching the institution ahead of time will better prepare a scholar for designing language in the experiment (be it a hypothetical scenario or questionnaire) that is clear, concise, and relevant to the elite's day-to-day work experience.

When carrying out elite experiments, having a clear organizational chart can be immensely helpful. This will allow one to target the specific agencies within an institution that are most relevant to the study. If an international institution does not have an organizational chart publicly available, contact the institution to request one, research existing scholarship to see if one can be found, and/or consult with experts on the institution to create one's own chart. Understanding the organization of the relevant organization is essential for accessing the correct sample, as well as having a sufficient understand of the constraints faced by the study population.

Considering elites' interests and constraints at the time of subject recruitment

Access to elites will be much improved if scholars are considerate of elites' interests and the constraints they face in participating in research. Gaining access remains one of the biggest challenges to conducting elite experiments. Yet it is critical that the sample is an accurate reflection of the type of decision-makers of interest to the scholar. If one's elite sample is not representative, it may lack external validity. In elite experiments, it is difficult to identify the composition of the sample, and then to gain access to them.

Identifying the correct sample can be difficult for several reasons. First, not all staff of government agencies or international organizations are transparently listed online. Second, even when names are identifiable, there is a great deal of turnover among staff, rendering information outdated and in need of thorough vetting. Third, even when names are available and correct, scholars need to obtain contact information of the individual elites. Thus, the acquisition of whom to contact and how requires a sustained effort.

One practical tip to identify contact information is to research the domain names for all relevant institutions (e.g. agencies, delegations, member states, etc.). One can then combine publicly available lists of elite names with knowledge of standard email domain

name formats (e.g. `firstname.lastname@hq.nato.int`) to contact participants. For each of us, combining this approach with publicly available data resulted in higher sample and response rates. Of course, personal contacts can be tremendously useful. However, even if direct contacts do not exist, reaching out for help to colleagues in the discipline who have worked with elites and who could establish important contacts can be a feasible strategy to start the process of costly elite sample recruitment.

Sample recruitment and the implementation of elite experiments can also be facilitated by conducting interviews with elites who work in the relevant field of study. These interviews can provide in-depth knowledge about the epistemic communities that scholars target and can be essential for gaining access. In our own research, we found that some elites would only agree to participate in paper or online experiments if we traveled to meet them in person. Since they often work with classified or at least private information and in secure buildings (e.g. parliaments, IO and NGO headquarters, etc.), elites have legitimate cybersecurity and identity concerns. They wonder who is actually on the other end of an email exchange or who would be monitoring any online responses to a survey (e.g. a foreign government, perhaps). With respect to identity concerns, one of the authors of this study arrived at many meetings with elites to find them holding printouts of the author's photo and CV. Such experiences reiterate the need for scholars to establish a professional online presence (e.g. university or personal website), preferably with a photo, before conducting fieldwork. Other ways to increase response rates and ensure a positive experience for both scholar and elite is to keep the survey instrument as brief as possible (i.e. five to ten minutes in length) and to couple it with an interview. Elites are often eager to verbally share their thoughts and experiences, and they are accustomed to such a format (Aberbach and Rockman, 2002).

The level of elite participation in experiments is likely associated with the types of experiments that scholars seek to employ in research. To date, the majority of experiments featuring elites are based on paper and pencil or online data collection methods. Neuroscience experiments as proposed by McDermott (2011), though promising for studying important questions about IR, may discourage participation by elites. Procedures such as brain imaging, hormonal analysis, and genetic mapping are not only time-consuming and difficult to administer, but are likely to be considered intrusive by elites and have lower rates of participation.

Framing research in ways that are both accurate and appropriate to the sample

The good news is that once the relevant sample is identified and contact information acquired, the willingness of potential respondents to participate in traditional paper and pen and online experiments appears high. It is true that elites are very busy (Hafner-Burton et al., 2013). However, this does not mean elites will not take the time to answer a survey or participate in an interview. Swedlund (2017a) had a 53% response rate to an online survey experiment, and over 90% of the elite respondents expressed a desire to be kept informed about the results of the research. Another survey experiment study had an 85% response rate, and, similarly, nearly all of the elite respondents indicated interest in the findings of the study (Hardt, 2018a). Elites are busy, but if you are asking relevant

and important questions to their work, they will find the time to participate in your research. In contrast, members of the general public are not likely to be as intrinsically motivated to participate.

When initially approaching the elite, we advise putting your request directly in the subject line of your e-mail: that is “Meeting or Interview Request” or “Request to participate in survey.” We also recommend that you mention any overlapping institutional affiliations or connections you might have (as long as you have permission to do so). Elites are often a tightly dense network. Once you have infiltrated this network, you should not be afraid to use it to your advantage with IRB permission. Finally, do not be put off by a lack of response, and do not be afraid to send your request multiple times. While you should always be polite with your requests and acknowledge that you are asking for a participant’s valuable time, you should also not assume that a non-response is a no. It is also important to make sure the relevance of the research is very clear. Doing so can get buy-in from elite respondents. It is best to avoid academic jargon or narrow theoretical concepts when approaching an elite—instead focusing on the big-picture implications of the research. Focusing on these also helps to avoid biasing your research since one must balance informed consent while not revealing too much information about what it is you are testing.

Research framing matters particularly at two crucial points: first, during subject recruitment (i.e. the initial request) and second, during informed consent (e.g. paper handout or a script read aloud by the scholar). The latter reiterates information in the initial request, tells the subject how the experiment will proceed, and invites the subject to say yes or no to participating. (Consult with your institution’s IRB for examples.) Here, we focus on the initial request since this can make or break response rates and since the experiment itself will vary widely depending on the research question. Most importantly, the written request (e.g. email, letter, fax) should be no longer than a paragraph to increase likelihood of receiving a response.

We suggest the following template. First, address the elite with his or her full current and formal title (e.g. Minister, Colonel, Dr., etc.). Civilian and military ranks can change in just a few years, so one needs to be sure that the title listed online is from a recent website. Second, one should introduce one’s self, citing name, academic status (e.g. researcher, professor, etc.), and institutional affiliation. If available, one should add here a sentence that helps with credibility, such as a reference to a shared professional contact, shared institutional history (e.g. alumnus/a of ABC university) or relevant professional experience (e.g. worked as an intern at XYZ organization). Third, one should state the request for a meeting, interview, or appointment and state what it would include (e.g. survey or other experimental instrument). It is also important to specify dates and times when one would be available rather than suggesting that one is perennially available. Fourth, one should provide a one-sentence description of the research project. Fifth, there should be a sentence explaining the purposes of the research and how it will be used (e.g. scholarly article, book, etc.). Sixth, the scholar should specify if and how he or she would like to use the elite’s name in publications or if and how the elite’s identity will be anonymized. Finally, the email should thank the elite and conclude with a signature giving the scholar’s full academic title, email address, personal/professional website or LinkedIn profile, Twitter handle, and a link to a recent publication if/where one is available.

Regarding persistence, one of the authors has a 3-day rule. If an elite (or an assistant) does not reply within 3 days, the author sends another email and continues to send an email every 3 days until the scholar either receives a reply or has sent a total of three emails. If a phone number is publicly available, the author then begins follow-up phone calls 4 days after the initial email interview request. Scholars should be prepared to, upon request from elites or their assistants, provide additional evidence of who the scholar is (e.g. copy of passport), provide sample questions from the interview and/or experiment, and provide any formal documentation (e.g. IRB approval, a grant award, etc.) from the scholar's university that indicates that the project is legitimate and registered with the university. In some cases, an elite may even ask for a reference, such as a supervisor or colleague at the university who his/her assistant can call to verify the scholar's identity. Members of the general public would not likely ask for these additional steps.

If one is using verbal informed consent, it should be a short and straightforward memorized script that one is prepared to say upon sitting down in an elite's office (or beginning the conversation over the phone). Many IRB offices have suggested scripts that scholars can employ and edit. The verbal informed consent involves the scholar saying out loud essentially everything that was in the interview request email so the elite is fully informed of the project and how his/her identity and information will be used. The only difference is that at the end, the scholar asks the elite if he or she may proceed with the interview and/or survey.

Providing elites the assurance of anonymity

When conducting research with elites, researchers need to make sure that the respondent knows that they will be kept anonymous, as there may be professional costs for participating in the research. Such risks will be substantially mitigated if anonymity is guaranteed and the respondent is sufficiently assured that their inputs will be securely recorded. Anonymity gives the respondent space to be forthcoming about the challenges both they and their organization face. At times, it may also be important to keep the respondent's agency anonymous. Although researchers prefer to be as transparent as possible about the study subjects and where they work, insisting on identifiable information could be costly insofar as it could result in significant delays, or it could bias the research. It may encourage respondents to answer the survey according to how they think their agency would *want* them to respond, rather than based on their actual experiences working for the agency.

Sending tailored follow-up emails

Finally, after the research has been conducted, we encourage scholars to follow up with respondents. Send your participants a thank you in which you offer to keep in touch and share your results. If a respondent asked to be kept informed about the results of the study, *actually do so*. Let them know about key publications. Offer to present your research for their colleagues. Be a point of contact for them about the scholarly findings in your field. One simple way to ensure that one can be in contact years later—particularly when scholarship may take a year or more to publish—is to submit a LinkedIn contact request to the elite after the meeting. Additionally, it is important to follow

through with any requirements of one's IRB office or Ethics Board. For example, some offices require that scholars send a follow-up email explaining the study in greater detail if any deception or omission of information was involved.

Following up with and maintaining contact with respondents is important to not only doing your due diligence as a scientist, but it also has important practical benefits for future research projects. Scholars may want to sample these individuals again in a future study or need their help to identify other relevant participants. As a scholar, one has likely invested a lot of time, and perhaps money, into identifying their sample. Do not simply abandon this knowledge when the project is complete. Instead, build on it to help streamline the process in the future.

Conclusion

In this article, we argued that foreign policy elites, as subjects, deserve closer inspection in IR, and that elite experiments offer scholars a unique opportunity to better understand their behavior and preferences. To specify when scholars should use the method, we have identified types of research that could most benefit from the method and other types that would not. Importantly, we provided a detailed set of guidelines for scholars who are considering using the method for the first time and for scholars who already use the method but are seeking ways to improve response rates, while still ensuring ethical considerations.

Our article justifies the use of elite experiments by noting that samples of specific elites who relate directly to one's research question are fundamentally different from other types of samples, such as convenience samples of elites or samples of non-elites (e.g. college students, etc.). That is, elites have disproportionate influence over foreign policy outcomes and experience unique structural constraints—meaning that studies involving targeted elite samples can lead different results than non-elite samples. We argue that the value of the method lies in several factors. Scholars can use it to better specify their theories. The method can facilitate theory-testing on a sample highly relevant to one's research question and can help overcome the social desirability bias. Using elite experiments can also reveal new hypotheses for subsequent studies and help bridge the academic–policy divide.

Moving forward, we recommend several directions for future research on elite experiments in IR. First, scholars interested in questions of the legitimacy of international institutions—from IOs to NGOs to international regimes—can conduct elite experiments to see how elites' opinions of their own institutions contrast with those of domestic publics. Are there gaps in perceptions of legitimacy? Second, scholars can also adopt a comparative approach by conducting elite experiments. How does elite decision-making vary across different international institutions under different structural constraints? Third, scholars can also employ elite experiments to test existing theories about how elites perceive different policy issues. Do elites' private preferences vary significantly from their public preferences? How does this variation affect policymaking? Fourth, scholars can assess elites' private preferences in relation to their foreign policy behavior—from UN voting behavior to military interventions to human rights records. Fifth, there exists substantial room for future scholarship on how an elite's identity (e.g. gender, race, ethnicity and sexual orientation) affects elite behavior in foreign policy (Anievas et al. 2014; Benedix and Jeong 2020).

Ultimately, elite experiments can provide scholars with an alternative means of exploring key research questions on elite samples that most closely approximate elite populations of interest. The approach, however, is not without drawbacks. In support of methodological pluralism, our article stresses that elite experiments complement existing, well-established methodological approaches and help us answer important questions for the field of IR—but they are more useful in some research domains than others. We call on more IR scholars to consider adopting the method as an additional useful strategy for studying and explaining elite behavior and preferences.

Acknowledgements

The authors contributed equally to this article, and they are grateful for comments on previous versions by Daniel Butler, James N. Druckman, Anne Holthoefler, Felix Bethke, Jordan Tama, Daniel L. Nielson, Hannah Smidt, D.G. Kim, Adam Scharpf, Felix Haas, and two anonymous reviewers.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Heidi Hardt  <https://orcid.org/0000-0001-5014-4975>

Haley J. Swedlund  <https://orcid.org/0000-0001-7700-1501>

Notes

1. See, for example, Hafner-Burton et al., 2017; McDermott, 2002a, 2011; Mintz et al., 2011. The behavioral turn has roots in psychology and subsequently behavioral economics (i.e. Cramer et al., 2005; DellaVigna, 2009; Loxton et al., 2011).
2. In American Politics, see Butler and Broockman 2011; Broockman 2013; Butler and Powell 2014; Butler and Dynes 2016; Butler et al. 2017; Grose et al. 2015, as examples. In Comparative Politics, see Calfano et al. 2014, Calfano and Oldmixon 2016, Smith 2016, Calfano et al. 2017, and Lierl 2018, as examples.
3. Quoted from introductory test to Bayram and Graham's (2017) survey experiment.
4. In their vignettes, the scholars manipulate whether students learn about agreement or disagreement between the US and other IO member states about policies of the organization and budget size.
5. In this case, an experiment on officials from the US Treasury (who lead US engagement in multilateral organizations) or the US State Department (e.g. in the department's Bureau of International Organization Affairs or US Mission to the UN) could have provided more external validity.
6. Recent work reiterates that substantive differences do exist in the behavior between elites and the public, although the magnitude of the gap may be overstated (Kertzer, 2020).
7. If a scholar has limited or no access to an IRB (Yanow and Schwartz-Shea, 2008), we recommend consulting other universities' IRBs and the appendix in Gerber and Green's book (2012).
8. An experiment may be eligible for IRB exempt status, which often does not require written consent.

References

- Abbott K, Genschel P, Snidal D, et al. (2015) *International Organizations as Orchestrators*. Cambridge, MA: Cambridge University Press.
- Aberbach J and Rockman B (2002) Conducting and coding elite interviews. *Political Science & Politics* 35(4): 673–676.
- Aberbach J, Chesney J and Rockman B (1975) Exploring elite political attitudes: some methodological lessons. *Political Methodology* 2(1): 1–27.
- Anievas A, Manchanda N and Shilliam R (2014) Race and Racism in International Relations: Confronting the Global Colour Line. Abingdon, UK: Routledge.
- APSA (2012) . Second Edition. *APSA Committee on Professional Ethics, Rights and Freedoms: 1–40*. <http://www.apsanet.org/portals/54/Files/Publications/APSAEthicsGuide2012.pdf>
- Avey P and Desch M (2014) What do policymakers want from us? Results of a survey of current and former senior national security decision makers. *International Studies Quarterly* 58(2): 227–246.
- Axelrod R (2015) *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, NJ: Princeton University Press.
- Baekgaard M and Serritzlew S (2016) Interpreting performance information: motivated reasoning or unbiased comprehension. *Public Administration Review*, 76: 73–82.
- Bartlett T (2014) Dartmouth and stanford apologize after a political-science experiment gone wrong. *Chronicle of Higher Education* October 29. <https://www.chronicle.com/article/DartmouthStanford/149687>
- Basken P (2015) Embrace of deception in experiments puts social scientists in an ethical bind. *Chronicle of Higher Education*. September 14. <https://www.chronicle.com/article/Embrace-of-Deception-in/233077>
- Bayram AB (2017) Due deference: cosmopolitan social identity and the psychology of legal obligation in international politics. *International Organization* 71(S2017): S137–S163.
- Bayram AB and Graham ER (2017) Financing the United Nations: explaining variation in how donors provide funding to the UN. *The Review of International Organizations* 12(3): 421–459.
- Bayram AB and Holmes M (2019) Feeling their pain: affective empathy and public preferences for foreign development aid. *European Journal of International Relations*. Online Edition. <https://journals.sagepub.com/doi/abs/10.1177/1354066119890915>
- Benedix W and Jeong GH (2020) Gender and Foreign Policy: Are Female Members of Congress More Dovish than their Male Colleagues? *Political Research Quarterly* 73(1): 126–140.
- Bennett A (2013) The mother of all -isms: causal mechanisms and structured pluralism in international relations theory. *European Journal of International Relations*. 19(3): 459–481.
- Bethke F (2016) Cultural Bias in the Perception of Global Events. Global Cooperation Research Papers, 14. Centre for Global Cooperation Research. https://www.gcr21.org/fileadmin/web-site/daten/pdf/Publications/Bethke_Cultural-Bias-Perception-Foreign-Policy_2198-0411-GCRP-14.pdf
- Blainey G (1988) *Causes of War*. New York: Free Press.
- Bloom M (2012) *Bombshell: Women and Terrorism*. Philadelphia, PA: University of Pennsylvania.
- Broockman D (2013) Black politicians are more intrinsically motivated to advance blacks' interests: a field experiment manipulating political incentives. *American Journal of Political Science* 57(3): 521–536.
- Busby J, Kafura C, Monten J, et al. (2019) Multilateralism and the use of force: experimental evidence on the views of foreign policy elites. *Foreign Policy Analysis* 16(1): 118–129.
- Butler D and Broockman D (2011) Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science* 55(3): 463–477.

- Butler D and Dynes A (2016) How politicians discount the opinions of constituents with whom they disagree. *American Journal of Political Science* 60(4): 975–989.
- Butler D and Powell E (2014) Understanding the party brand: experimental evidence on the role of valence. *Journal of Politics* 76(2): 492–505.
- Butler D, Volden C, Dynes A, et al. (2017) Ideology, learning, and policy diffusion: experimental evidence. *American Journal of Political Science* 61(1): 37–49.
- Calfano B and Oldmixon E (2016) Remembering to ask the noss: priming and the dynamics of priest reliance on bishop cues. *Religions* 7(3): 21.
- Calfano B, Oldmixon E and Gray M (2014) Strategically prophetic priests: an analysis of competing principal influence on clergy political action. *Review of Religious Research* 56(1): 1–21.
- Calfano B, Michelson M and Oldmixon E (2017) *A Matter of Discretion: The Politics of Catholic Priests in the United States and Ireland*. Lanham, MD: Rowman & Littlefield.
- Caprioli M and Boyer MA (2001) Gender, violence, and international crisis. *Journal of Conflict Resolution* 45(4): 503–518.
- Checkel J (2005) International institutions and socialization in Europe: introduction and framework. *International Organization* 59(4): 801–826.
- Coffman L and Niederle M (2015) Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives* 29(3): 81–98.
- Costa M (2017) How responsive are political elites? A meta-analysis of experiments on public officials. *Journal of Experimental Political Science* 4(3): 241–254.
- Cramer C, Loewenstein G and Prelec D (2005) Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature* 43(1): 9–64.
- Davis J and McDermott R (2020) The past, present and future of behavioral IR. *International Organization* 1–31. Online Edition.
- DellaVigna S (2009) Psychology and economics: evidence from the field. *Journal of Economic Literature* 47(2): 315–372.
- Dietrich S (2016) Donor political economies and the pursuit of aid effectiveness. *International Organization* 70(1): 65–102.
- Dietrich S (2021) *States, Markets, and Foreign Aid*. Cambridge, MA: Cambridge University Press.
- Druckman J and Kam C (2011) Students as experimental participants: a defense of the ‘Narrow Data Base’. In: Druckman J, Green D, Kuklinski J, et al. (eds) *Cambridge Handbook of Experimental Political Science*. Cambridge, MA: Cambridge University Press.
- Dunne T, Hansen L and Wight C (2013) The end of international relations theory? *European Journal of International Relations* 19(3): 405–425.
- Eichenberg R (2016) Public Opinion on Foreign Policy Issues. Oxford Research Encyclopedia of Politics. Retrieved 14 January 2021, from <https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-78>.
- Fehr E and List JA (2004) The hidden costs of returns of incentives—trust and trustworthiness among Ceos. *Journal of the European Economic Association* 2(5): 743–771.
- Findley M, Nielson D and Sharman J (2013) Using field experiments in international relations: a randomized study of anonymous incorporation. *International Organization* 67(4): 657–693.
- Findley MG, Harris AS, Milner HV, et al. (2017a) Who controls foreign aid? Elite versus public perceptions of donor influence in aid-dependent Uganda. *International Organization* 71(4): 633–663.
- Findley MG, Milner HV and Nielson DL (2017b) The choice among aid donors: the effects of multilateral vs. bilateral aid on recipient behavioral support. *Review of International Organizations* 12(2): 307–334.

- Finnemore M (1993) International organizations as teachers of norms: the United Nations educational, scientific, and cultural organization and science policy. *International Organization* 47(4): 565–597.
- Fortna VP (2015) Do Terrorists Win? The Use of Terrorism and Civil War Outcomes 1989–2009. *International Organization* 69(3): 519–556.
- George A (1993) *Bridging the Gap: Theory and Practice in Foreign Policy*. Washington, D.C.: United States of Peace Press.
- Grohs S, Adam C and Knill C (2016) Are some citizens more equal than others? Evidence from a field experiment. *Public Administration Review* 76(1): 155–164.
- Grose C, Malhotra N and Van Houweling RP (2015) Explaining explanations: how legislators explain their policy positions and how citizens react. *American Journal of Political Science* 59(3): 724–743.
- Hafner-Burton E, Hughes DA and Victor D (2013) The cognitive revolution and the political psychology of elite decision making. *Perspectives on Politics* 11(02): 368–386.
- Hafner-Burton E, Haggard S, Lake D and Victor D (2017) The behavioral revolution and international relations. *International Organization* 71(S1): S1–S31.
- Hafner-Burton E, LeVeck B and Victor D (2017) No false promises: how the prospect of non-compliance affects elite preferences for international cooperation. *International Studies Quarterly* 61(1): 136–149.
- Hardt H (2014) *Time to React: The Efficiency of International Organizations in Crisis Response*. New York: Oxford University Press.
- Hardt H (2016) From states to secretariats: delegation in the African Union Peace and Security Council. *African Security* 9(3): 161–187.
- Hardt H (2018a) Who matters for memory: sources of institutional memory in international organization crisis management. *Review of International Organizations* 13(3): 457–482.
- Hardt H (2018b) *NATO's Lessons in Crisis: Institutional Memory in International Organizations*. New York: Oxford University Press.
- Hawkins D, Lake D, Nielson D, et al. (2006) *Delegation and Agency in International Organizations*. Cambridge, UK: Cambridge University Press.
- Hooghe L and Marks G (2014) Delegation and pooling in international organizations. *Review of International Studies* 10(3): 305–328.
- Hyde S (2010) The future of field experiments in international relations. *The ANNALS of the American Academy of Political and Social Science* 628(1): 72–84.
- Hyde S (2015) Experiments in international relations: lab, survey, and field. *Annual Review of Political Science* 18(May): 403–424.
- IO (2017) The behavioral revolution and international relations. *International Organization* 71(S1): S1–S277.
- ISQ (2011) Symposium on experimental research in international relations. *International Studies Quarterly* 55(2): 493–583.
- ISR (2007) Forum on behavioral IR. *International Studies Review* 9(1): 156–172.
- Jalalzai F (2008) Women rule: Shattering the executive glass ceiling. *Politics & Gender* 4(2): 205–231.
- Jervis R (1976) *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
- Jilke S, Van de Walle S and Kim S (2016) Generating usable knowledge through an experimental approach to public administration. *Public Administration Review* 76(1): 69–72.
- Kern F and Gleditsch K (2017) Exploring pre-registration and pre-analysis plans for qualitative inference. Working Paper. https://www.researchgate.net/profile/Florian_Kern4/publication/319141144_Exploring_Pre-registration_and_Pre-analysis_

- Plans_for_Qualitative_Inference/links/599455d60f7e9b98953af045/Exploring-Pre-registration-and-Pre-analysis-Plans-for-Qualitative-Inference.pdf
- Kertzer J and Renshon J (2015) Putting things in Perspective: Mental Stimulation in Experimental Political Science. Working Paper. https://people.fas.harvard.edu/~jkertzer/Research_files/PT%20Web%20version.pdf
- Kertzer J (2017) Resolve, time, and risk. *International Organization* 71(S1): S109–S136.
- Kertzer J (2020) Re-assessing elite-public gaps in political behavior *American Journal of Political Science*. Online Edition.
- Lavrakas PJ (2014) List-experiment technique. In: Lavrakas PJ (ed.), *Encyclopedia of Survey Research Methods*, New York: Sage Publishing, 433–435.
- Levitt SD and List JA (2007) What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives* 21(2): 153–174.
- Lierl M (2018) Corruption and Accountability at the Grassroots Level: An Experiment on the Preferences and Incentives of Village Leaders. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2727329.
- Lin W and Green D (2016) Standard operating procedures: a safety net for pre-analysis plans. *PS: Political Science & Politics* 49(3): 495–500.
- Loewen PJ, Rubenson D and Wantchekon L (2010) Help me help you: conducting field experiments with political elites. *The ANNALS of the American Academy of Political and Social Science* 628(1): 165–175.
- Lopez A, McDermott R and Peterson M (2011) States in mind: evolution, coalitional psychology, and international politics. *International Security* 36(2): 48–54.
- McDermott R (2002a) Experimental methods in political science. *Annual Review of Political Science* 5(1): 31–61.
- McDermott R (2002b) Experimental methodology in political science. *Political Analysis* 10(4): 325–342.
- McDermott R (2011) New directions for experimental work in international relations. *International Studies Quarterly* 55(2): 503–520.
- McNutt M (2015) Editorial retraction. *Science* 348(6239): 1100.
- Mintz A, Geva N, Redd S, et al. (1997) The effect of dynamic and static choice sets on political decision making: an analysis using the decision board platform. *American Political Science Review* 91(3): 553–566.
- Mintz A, Redd S and Vedlitz A (2006) Can we generalize from student experiments to the real world in political science, military affairs, and international relations? *Journal of Conflict Resolution* 50(5): 757–776.
- Mintz A, Yang Y and McDermott R (2011) Experimental approaches to international relations. *International Studies Quarterly* 55(2): 493–501.
- Olken B (2015) Promises and perils of pre-analysis plans. *Journal of Economic Perspectives* 29(3): 61–80.
- Pakulski J (2008) Elite theory. In: Darity Jr. WA (ed.), *International Encyclopedia of the Social Sciences*. Vol. 2, 2nd edn. Reference USA: Macmillan.
- Pareto V (1935) *The Mind and Society*, Livingstone A (ed.), Vol. 4, New York: Harcourt Brace.
- Peabody R (1990) Interviewing political elites. *PS: Political Science & Politics* 23(3): 451–455.
- Redd S (2002) The influence of advisers on foreign policy decision making an experimental study. *Journal of Conflict Resolution* 46(3): 335–364.
- Renshon J (2015) Losing face and sinking costs: experimental evidence on the judgment of political and military leaders. *International Organization* 69(3): 659–695.
- Reynolds A (1999) *Electoral Systems and Democratization in Southern Africa*. New York: Oxford University Press.

- Risse T and Sikkink K (1999) The Socialization of International Human Rights Norms into Domestic Practices: Introduction. In: Risse T, Sikkink K and Ropp SC (eds) *The Power of Human Rights: International Norms and Domestic Change*. Cambridge, UK: Cambridge University Press. 1–38.
- Sasley BE (2010) Affective attachments and foreign policy: Israel and the 1993 Oslo Accords. *European Journal of International Relations* 16(4): 687–709.
- Sandholtz W and Stiles K (2009) *International Norms and Cycles of Change*. New York: Oxford University Press.
- Saunders EN (2018) Leaders, advisers, and the political origins of elite support for war. *Journal of Conflict Resolution* 62(10): 2118–2149.
- Schneider C (2015) Mind the Gap? Links Between Policy and Academic Research in Foreign Aid.” Unpublished manuscript.
- Sheffer L, Loewen PJ, Soroka S, et al. (2018) Nonrepresentative representatives: an experimental study of the decision making of elected politicians. *American Political Science Review* 112(2): 302–321.
- Simmons B (2009) *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge, UK: Cambridge University Press.
- Smith AE (2016) When clergy are threatened: Catholic and Protestant leaders and political activism in Brazil. *Politics and Religion* 9(3): 431–455.
- Stouffer S (1956) *Conformity, and Civil Liberties: A Cross-Section of the Nation Speaks Its Mind*. Cambridge, MA: Cambridge University Press.
- Sullivan J, Walsh P, Shamir M, et al. (1993) Why politicians are more tolerant: selective recruitment and socialization among political elites in Britain, Israel, New Zealand and the United States. *British Journal of Political Science* 23(1): 51–76.
- Swedlund H (2017a) Can foreign aid donors credibly threaten to suspend aid? Evidence from a cross-national survey of donor officials. *Review of International Political Economy* 24(3): 454–496.
- Swedlund H (2017b) *The Development Dance: How Donors and Recipients Negotiate the Delivery of Foreign Aid*. Ithaca, NY: Cornell University Press.
- Tansey O (2007) Process tracing and elite interviewing: a case for non-probability sampling. *PS: Political Science and Politics* 4(4): 765–772.
- Tetlock P (2005) *Expert Political Judgment: How Good is it? How Should we Know?* Princeton, NJ: Princeton University Press.
- Tomz M, Weeks J and Yarhi-Milo K (2020) Public opinion and decisions about military force in democracies. *International Organization* 74(1): 119–143.
- Yarhi-Milo K, Kertzer J and Renshon J (2018) Tying hands, sinking costs, and leader attributes. *Journal of Conflict Resolution* 62(10): 2150–2179.

Author biographies

Simone Dietrich is an Associate Professor of Political Science and International Relations at the University of Geneva in Switzerland. Her research answers questions in the fields of international and comparative political economy. Her book *States, Markets, and Foreign Aid* is forthcoming with Cambridge University Press and her research appears in journals such as *International Organization*, *Journal of Politics*, *International Studies Quarterly*, *Journal of Experimental Political Science*, *Review of International Organizations*, *World Development*, among others.

Heidi Hardt is an Associate Professor of Political Science at the University of California, Irvine and the author of two books: *NATO's Lessons in Crisis: Institutional Memory in International Organization* and *Time to React: The Efficiency of International Organizations in Crisis Response*.

Her research examines the impacts of international organizations on international peace and security, and her work appears in journals such as *Journal of Politics*, *Global Governance*, *Review of International Organizations*, *PS: Political Science & Politics*, *Journal of Global Security Studies*, and *European Security and African Security*.

Haley J. Swedlund is an Assistant Professor of International Relations at the Department of Political Science, Centre for International Conflict Analysis and Management (CICAM) at Radboud University. Her research SITS at the nexus of international political economy, development, and conflict studies. She is the author of *The Development Dance: How Donors and Recipients Negotiate the Delivery of Foreign Aid* (Cornell University Press, 2017), and her research has appeared in journals such as *African Affairs*, *Review of International Political Economy*, *International Affairs*, and *Social Forces*.