Experiments and Surveys on Political Elites

Joshua D. Kertzer,[1] Jonathan Renshon[2]

[1]Department of Government, Harvard University, Cambridge, MA, 02138; email: jkertzer@gov.harvard.edu
[2]Department of Political Science, University of Wisconsin-Madison, Madison, WI, 53706; email: renshon@wisc.edu

Ann. Rev. Pol. Sci. 2022. 25:1–26

https://doi.org/10.1146/XXXX

Keywords

political elites, elite experiments, elite surveys

Abstract

One of the major developments in political science in the past decade has been the rise of experiments and surveys on political elites. Yet an acceleration in the number of elite studies has outpaced our collective understanding of best practices, and how we know a good elite experiment when we see one. In this article, we discuss some of the challenges in the study of political elites – from who counts as an elite, to how to best utilize elite experiments in the context of broader research designs. We also offer a number of recommendations for questions of access, recruitment, and representativeness, as well as designs researchers can use to study eliteness without access to elites.

Contents

1. INTRODUCTION

One of the major developments in political science in the past decade has been the rise of experiments and surveys on political elites. This is perhaps unsurprising given that many of our theories of politics are either directly or indirectly about the beliefs or behavior of elites, whether at the local, national or international level. The increased use of elite samples has occurred in all of the top journals in the discipline, in all of the empirical subfields of political science: from American Politics (e.g., Broockman & Skovron 2018, Hertel-Fernandez et al. 2019) to Comparative Politics (e.g., Grossman & Michelitch 2018, Sheffer et al. 2018, Pereira 2021) to International Relations (IR) (e.g., Hafner-Burton et al. 2014, Findley et al. 2017a). It is also evident in terms of institutional practices: elite experiments have been the subject of dedicated mini-conferences at the Midwest and Southern Political Science Association annual meetings, and at a wide range of universities.

While the idea of studying elites directly as survey respondents or experimental subjects dates to a much earlier period in political science (for early examples, see Miller & Stokes 1963, Putnam et al. 1976, Oldendick & Bardes 1982), recent shifts towards microfoundations, causal identification, the rise of survey and field experiments, and the rise of interest in leaders more generally have greatly increased the use of elite subjects in political science studies. Americanists have turned to samples of policymakers to study spending priorities (Arceneaux et al. 2018), legislators' use of public opinion (Butler & Nickerson 2011), and responses to fact-checking (Nyhan & Reifler 2015). Comparativists have used survey experiments on elites to study motivated reasoning (Baekgaard et al. 2019), risk-propensity (Heß et al. 2018, Sheffer & Loewen 2019, Linde & Vis 2017) and issue framing (Walgrave et al. Forthcoming) in developed and developing countries alike. In IR, scholars have examined questions using elite samples ranging from perceptions of costly signals (Yarhi-Milo et al. 2018) to obligation to international law (Bayram 2017), to attitudes towards the use of force (Tomz et al. 2020). A similar surge has taken place outside of political science, most notably by economists interested in questions of expertise and leadership (e.g., Potters & van Winden 2000, Palacios-Huerta & Volij 2009, List & Mason 2011, Banuri et al. 2019) and by psychologists interested in dominance, hierarchy and power (Sherman et al. 2016, van der Meij et al. 2016).

In this article, we step back and take stock of the promises and pitfalls of these approaches. Given space constraints and the number of excellent recent review articles that have been written on a variety of types of elite studies, including field experiments on political institutions (Grose 2014), audit studies (Costa 2017), and survey and interview methods on elites (Rivera et al. 2002, Hoffmann-Lange 2007, 2008, Rodríguez-Teruel & Daloz 2018), our predominant focus here is on elite experiments, defined here as studies where a sample of political elites has been randomly assigned by the experimenter to treatment conditions. In this sense, we bracket a vibrant literature utilizing elite surveys for largely descriptive purposes — for example analyzing the policy preferences of the top one-percent (Page et al. 2013), political scientists' perceptions of democratic backsliding (Carey et al. 2019)), or simulations like wargames that don't contain experimental manipulations (Lin-Greenberg et al. 2020) — to focus on experimental studies where researchers are interested in using elite samples to make causal inferences. Given that many of the unique logistical challenges associated with studying political elites include issues of recruitment and access, our discussion is particularly focused on types of experiments in which respondents are aware they are being studied, thereby setting aside audit or field experiments in which elites may not be aware they are study subjects, which are addressed in other recent reviews (e.g., Butler & Crabtree 2021). However, many of our points below apply to elite experiments and surveys more generally.

Elite experiments: Studies where a sample of political elites has been randomly assigned to treatment conditions

We begin by offering several observations based on a quantitative literature search of elite experiments in the "big 3" journals in Political Science over the last twenty years, fleshing out those developments with a broader discussion of some of the challenges that have arisen in this literature thus far. We suggest that elite studies are the most useful when they test theories that directly implicate elites' domain-specific expertise and experience, but that the use of elite samples in and of themselves does not resolve concerns about generalizability. Second, we present a framework for thinking through conceptual issues relating to elite experiments, enumerating three different conceptions of eliteness—occupational, compositional, and cognitive—and providing a checklist for scholars to use to answer relevant questions about how their theory maps on to their sample. We use these distinctions to suggest when elite studies are particularly valuable, proposing a number of ways that scholars can integrate elite studies into their broader research designs. In particular, we advocate for what we call "complementary designs", in which scholars field studies on both mass and elite samples, but using each sample to test a different component of the theory. Finally, we offer a practical discussion of how to study elites, concerning questions of how elite respondents can be recruited, how to think about questions of representativeness, and how to experimentally study elites when researchers might lack access to elites themselves. Baking "eliteness" into experimental designs themselves not only lowers barriers to entry in the study of political elites but also provides additional causal leverage on the qualities that make elites unique in the first place.

2. THE RISE OF ELITE EXPERIMENTS

For a long time, elites and masses in political science were studied using very different methods (Kertzer & Tingley 2018). The rise of opinion polls in the 1930s (Moyser & Wagstaffe 1987, 5), coupled with the rise of institutionalist approaches (Ricart-Huguet 2019), caused interest to wane in the study of political elites, precisely out of an assumption that survey methods could be more fruitfully applied to mass rather than elite political

behavior — despite entreaties that some elites were more accessible than others (e.g., state legislators, Maestas et al. 2003). Instead, political scientists leveraged virtually every other method and source of data to study elites, including speeches, diaries, and autobiographies (George & George 1964), roll-call votes (Poole & Rosenthal 1997), cognitive maps and operational codes (George 1969, Axelrod 1976), participant observation among lawmakers and diplomats (Fenno 1978), and open-ended or semi-structured interviews (Zuckerman 1972). This has continued today, as scholars continue to productively use non-survey or experimental approaches to the study of leaders, including text-based methods to study elite personality at a distance (Ramey et al. 2016), archival methods (Saunders 2011, Yarhi-Milo 2018), biographical approaches (Goemans et al. 2009, Fuhrmann & Horowitz 2015, Krcmaric et al. 2020), ethnographic approaches (Neumann 2012, Bussell 2020, Nair 2021), and network methods (Keller 2016, Mahdavi et al. 2017).

One notable trend over the past several decades, however, has been the rise of elite surveys and experiments. To characterize trends in elite experiments in political science, we conducted a quantitative literature review, manually compiling information on all experimental articles published in the "big 3" journals in political science (the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics*) from 2000-2020. A total of 914 experiments were published in 501 articles in these three journals over this 20 year period; 76 of the experiments were fielded on elite samples. **Figure 1** shows that while the number of experiments on elite samples published in these journals is dwarfed by the colossal increase in number of experiments published on non-elite samples (nationally representative samples of mass publics, convenience samples recruited through Amazon Mechanical Turk, student samples, etc.), there was a noticeable increase in the number of elite experiments beginning in 2015. **Figure 2** shows that elite experiments published in these three journals tend to differ from experiments conducted on non-elite samples in a variety of ways: elite experiments tend to feature significantly smaller sample sizes, for example, and are less disproportionately focused on American politics. **Figure 3** shows that elite experiments tend to feature a variety of experimental approaches: although a plurality of both elite and non-elite experiments are survey experiments, a higher proportion of the elite experiments were field and audit experiments than non-elite experiments were (unsurprising given what a non-elite audit study would look like). Reflecting the difficulty of hauling elites into the lab, none of the lab experiments published in these three journals during this time period were conducted on elites – though lab experiments on elites were published in other journals in this time period (e.g., Butler & Kousser 2015, Renshon 2015).

Perhaps unsurprisingly, the elite studies vary widely in terms of the types of elites recruited: from California school board members (Flavin & Hartney 2017) to international corporate service providers (Findley et al. 2017b), to activists at environmental NGOs (Hafner-Burton et al. 2016), to political scientists and historians (Tetlock & Lebow 2001), to Vietnamese legislators (Malesky et al. 2012). While much of the literature on elite experiments in political science remains fairly American-centric, only 43% of the elite experiments published in these three journals were fielded on political elites in the United States.

There are a number of other important differences as well in terms of how the studies were fielded. One is in terms of compensation: whereas 93% of the experiments on mass samples reported providing compensation to participants (or were conducted on platforms where compensation can be implied), only 3% of the experiments on elite samples mentioned compensating participants for their time. Another is in terms of recruitment. As **Figure 3**

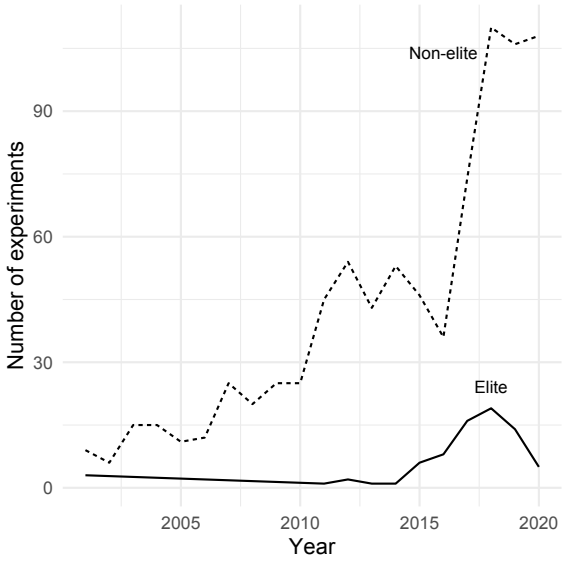

Figure 1: An analysis of experiments published in AJPS, APSR and JOP from 2000-2020 illustrates the growth in elite experiments over time. Our unit of analysis here is the experiment rather than the article.

shows, the modal experiment on a mass sample recruited respondents through online survey panels (51%), whereas the modal experiment on an elite sample recruited respondents via email solicitation (50%). For both mass and elite experiments, experimenters also often partner with local organizations for fielding experiments on both types of samples – particularly field experiments on elites.

While most of the articles included studies fielded just on mass or elite samples, 20% of the articles that included experiments on elite samples also included experiments on mass samples, often times to test how the two groups differ from one another (Kertzer 2021). The elite experiments also vary in the extent to which they focus on questions related to elites' domain-specific expertise. Many of the elite experiments in the data test theories related to the specific elites being studied, such as Broockman et al. (2019), who use a sample of technology entrepreneurs to show how Silicon Valley is more conservative on issues of government regulation than their other political attitudes might predict. Other experiments use elite samples to study more general questions about elite cognition, such as Sheffer et al. (2018), who test whether common biases from the judgment and decision-making literature hold in samples of national and provincial politicians (see also Linde & Vis 2017).

Given the extent to which elites are typically thought of as a hard to access population (e.g., Hafner-Burton et al. 2013), it is striking that elite response rates reported are often relatively high: 89% of the elite experiments provide information about their response rate, with the mean response rate reported as 40%. While this is lower than the average response rate reported for the mass experiments in the sample (55%), only 25% of mass experiments in the sample report a response rate, particularly because response rates for survey experiments fielded online are often unknown or poorly defined, and response rates for surveys conducted over the telephone are relatively low; Kennedy & Hartig (2019) reported in 2019

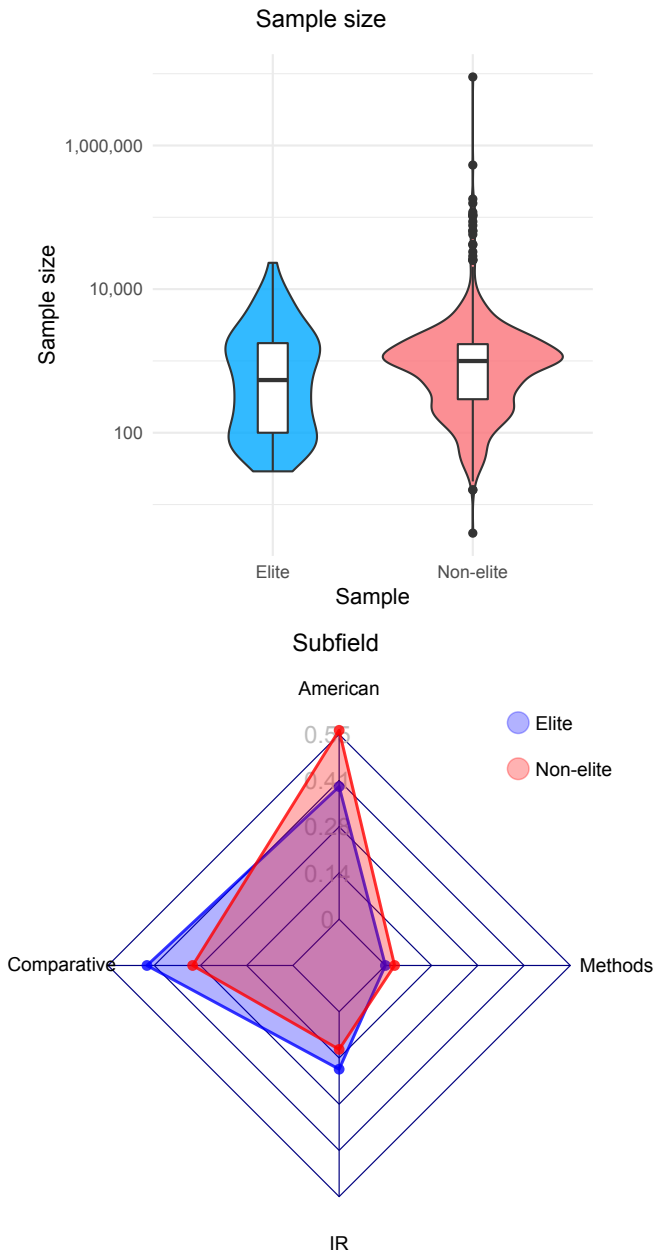

Figure 2: Differences between elite and non-elite experiments published in AJPS, APSR, and JOP from 2000-2020 in terms of sample size and subfield. Elite experiments in blue, non-elite experiments in red.

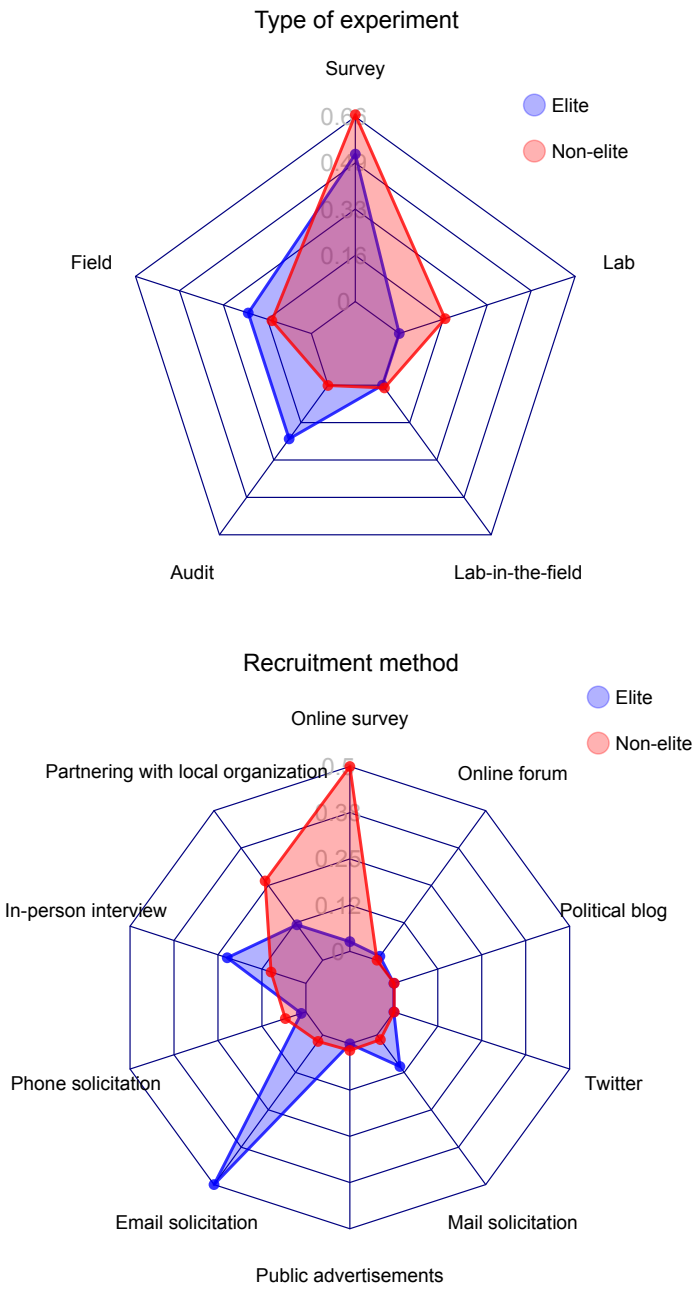

Figure 3: Differences between elite and non-elite experiments published in AJPS, APSR, and JOP from 2000-2020 in terms of type of experiment and recruitment method. Elite experiments in blue, non-elite experiments in red.

that the average response rate in Pew's telephone surveys was 6%. Importantly, there is also considerable heterogeneity in response rates across different types of elite experiments: field experiments on political elites report a relatively high response or completion rate (an average of 65%), whereas survey experiments on political elites report a much lower one (an average response rate of 28%; for survey experiments on American political elites, the average response rate is only 15%, and is generally lower in surveys of federal elites than municipal elites).

3. CHALLENGES IN THE STUDY OF POLITICAL ELITES

Overall, elite experiments in political science have yielded a number of substantively significant empirical and theoretical advances. And there is no doubt that elite experiments have benefited from broader trends in the field, as scholars have taken advantage of increasingly sophisticated experimental designs and reaped the benefits of more advanced methodological training at the graduate level. However, an acceleration in the use of elites in political science has not been accompanied by sufficient attention to questions about how elite respondents most productively fit into our research designs, or how to use the study of elites to best advance knowledge in the field. This neglect has led to a number of issues regarding how to most efficiently design and learn from elite experiments.

First, who counts as a political elite in the first place? Political scientists have used the term to refer to a wide range of actors, from academics, think tankers and local elites to presidents and prime ministers. There is considerable variation in the term both within and — especially — across subfield boundaries, with these different terminological boundaries reflecting very different conceptions of "eliteness," all of which have implications for how to most appropriately use experiments in service of testing our theories of politics. Compounding these issues is a tendency to elide discussion of how well our elite samples fit our theories.

Second, a review of the existing literature suggests some uncertainty about how elite subjects fit into broader research designs and, in particular, how to use them most efficiently in the service of theory-testing and probing generalizability. Even in the nascent literature on elite experiments, there is tremendous variation along these dimensions, with designs that use paired experiments fielded simultaneously on both elites and the mass public (Renshon 2015, Renshon et al. 2020), comparisons between different elite samples (Sheffer et al. 2018), designs that use elites for conceptual replications (Kertzer et al. 2021c), to test different sets of microfoundations (Tomz et al. 2020) or as part of a larger design that also incorporates observational data (Hertel-Fernandez et al. 2019). Up for debate is thus whether elite experiments are sufficient for theory-testing on their own given the limitations in their sample sizes and designs, when they are worth the additional logistical effort involved in fielding them, whether they are more profitably thought of as part of larger research designs, and what those designs might look like.

A third set of challenges involves the logistics of access and design in elite experiments, particularly access to elite populations and guidance on design considerations that might be especially pertinent to this specialized context. While there is already a cottage industry in providing advice along some of these dimensions, it focuses mostly on access itself rather than experimental design considerations (a category that includes discussion of how these studies fit into broader research agendas). After all, simpler, smaller experiments can be more compelling than larger, more complex ones if they provide tests that more closely

match the theory, or better complement the other empirical components of a project, and even a sample of very high level elites cannot make up for a design that doesn't accurately simulate the decision making process it's designed to illuminate. Moreover, the "how-to" guides available are often tailored towards relatively narrow subpopulations of elites and subfields of political science, leading to unnecessary confusion when the advice is aggregated for a broader audience. An example concerns the issue of compensation, which depending on what one reads is either critical to increasing participation rates, insulting to potential elite participants or not particularly effective in the first place.

In the discussion below, we discuss each challenge in turn.

4. WHO COUNTS AS AN ELITE?

One challenge in the study of political elites is widespread disagreement about who counts as a political elite in the first place (Putnam et al. 1976, Moyser & Wagstaffe 1987). Political scientists have been flexible in their use of the terminology, using the term to refer to anyone from business executives (Teigen & Karlsen 2019), to military officers (Lin-Greenberg 2021), to professors with doctorates in economics (Fatas et al. 2007), to mixed groups of government and military personnel recruited from executive education programs (Carnevale et al. 2011, Renshon 2015) and elected politicians (Kertzer et al. 2021c).

These disagreements are critical since the conception of eliteness one holds has implications for how elite surveys or experiments should be designed as well as how we aggregate findings across studies. If we define eliteness in terms of power, military generals would count as elites, but military cadets might not (Jost et al. 2017), whereas if we define eliteness in terms of domain-specific expertise, International Relations scholars would count as elites in the context of foreign policy (Busby et al. 2020a), but city councilors might not (Baekgaard et al. 2019). And what one subfield agrees on another would find less relevant; "local elites" — for example the slum leaders in India studied by Auerbach & Thachil (2020) — are relevant in studies of American and Comparative politics but would usually be less meaningful in the context of IR.

There are roughly three different ways in which elites have been conceptualized in the existing literature. The first are *occupational* approaches, which conceptualize political elites as actors whose institutional roles afford them higher levels of influence over public policy: prime ministers, legislators, civil servants, bureaucrats, diplomats, military officers, and so on (Putnam et al. 1976, Bussell 2020). Lasswell (1952, 6), for example, defines elites as "the holders of high positions in a given society" and political elites as "the power holders of a body politic" (p. 13). In this view, it is less that elites have unique traits, and more that they happen to occupy unique roles. Second are *compositional* approaches, which conceptualize elites as a distinctive class (Meisel 1962) identified via demographic categories, and focus on the social and economic characteristics that define them. Political elites tend to be wealthier, more educated, older, and more male than the population at large; in Western contexts, they are also more likely to be white (Carnes & Lupu 2015, Bahador et al. 2019, Gerring et al. 2019). Third are *cognitive* approaches, which emphasize the extent to which elites possess a distinctive constellation of cognitive traits, sometimes conceptualized as domain-specific expertise or experience (Hafner-Burton et al. 2013, Saunders 2017). Carnevale et al. (2011), for example, identify leaders as higher in decision-making competence compared to matched controls, while Dal Bó et al. (2017) show that elected politicians often have a different set of personality characteristics than the public as a whole. In many of these

Occupational models of elites: Political elites as actors whose institutional roles give them influence over public policy

Compositional models of elites: Political elites as a distinctive class characterized by common socioeconomic characteristics

Cognitive models of elites: Political elites as characterized by domain-specific expertise and experience

studies, it is not just that elites happen to possess traits like domain specific expertise, but rather, that domain specific expertise is seen as *constitutive* of what it means to be an elite (e.g., Hafner-Burton et al. 2013).

Three points are worth noting here. First, these varying formulations of eliteness are conceptually intertwined. Theories of elite cues in the study of political behavior, for example, operationalize political elites occupationally in terms of "persons who devote themselves full time to some aspect of politics or public affairs", which "include politicians, higher-level government officials, journalists, some activists, and many kinds of experts and policy specialists" (Zaller 1992, 6), yet it is because of their presumed domain-specific expertise that their voices are seen as carrying such weight (Lupia & McCubbins 1998). Similarly, studies of political selection and latent political ambition (Dynes et al. 2021, Gulzar 2021) often focus on the interplay of compositional and cognitive factors in exploring who decides to run for office, and whose campaigns are ultimately successful. Second, these conceptions are related in a more practical sense: samples of elites are often "bundled" in a manner that combines across these categories. This is particularly true for long-running elite panels — such as Chicago Council on Global Affairs Leader Survey, which is comprised of respondents from the executive and legislative branches, as well as foreign policy think tanks, labor and religious leaders, business officials and academics (Busby et al. 2020a,b) — but also common in respondent pools compiled by scholars. To illustrate the wide variation in who gets considered an elite, consider the finding from Renshon (2015), who shows — using a sample drawn from an executive education program at the Harvard Kennedy School along with a matched sample of controls — that the two groups overlap significantly on both subjective feelings and objective measures of power. Reassuringly, the leaders were higher on average on both measures, but the overlapping distributions suggests that at least some samples may not be as elite as we think while — depending on what aspect of eliteness we care about — more traditional samples of non-elites might be sometimes be just as appropriate.

Finally, these divergent conceptualizations probably reflect two factors: variation in the reasons scholars study elites in the first place as well as practical constraints. In some cases, we survey or conduct experiments on elites in order to test theories in which political elites are the relevant actors (Humayun 2021). In other cases, our theories are more general (particularly in work building on theories from social and cognitive psychology), and our interest is in examining whether elites also act or think in the manner suggested by the theory (Sheffer et al. 2018). Other studies focus instead on identifying the effect of eliteness on behavior, often using paired samples of political elites and ordinary citizens (Yarhi-Milo et al. 2018). Additionally, exigencies of sample size often motivate bundling of different types of elites (with different levels of eliteness), to mitigate concerns about statistical power and response rates. Similarly, although all three conceptualizations of elites are valuable, occupational and compositional conceptions of eliteness describe actors that are easier for researchers to recruit (e.g., highly educated) or simulate by incorporating aspects of elite environments into the research design, as discussed in detail below.

The rather wide variation in who gets considered elite has practical consequences. On one hand, broader conceptions of eliteness permit larger samples, more statistical power and the ability to make our designs more elaborate than they would otherwise be. It also increases the rate at which experiments may be fielded and published, accelerating the aggregation of knowledge across research areas. However, in the absence of clear criteria defining "elites," and clear expectations of the purpose served by elite samples in any given study, these highly heterogenous samples can sometimes prompt unnecessary confusion.

A first step is thus providing a clear conception of what "eliteness" is for researchers, but extant definitions are either helpfully narrow but subfield-specific, or general enough that they provide little practical guidance for most researchers. We suggest that the most productive way forward is less in a one-size-fits-all definition of eliteness, but rather, a checklist of what scholars conducting research on political elites should specify in their work.

Scholars working on political elites should:

1. Specify who the relevant political elites are for the question they are studying.
2. Indicate the dimensions on which their respondents can be conceptualized as elites, and
3. Explicitly assess the level of fit between the elites that are the focus of their theory and the elites that are studied in their empirics.
4. In the case of bundled samples, we advise as transparent a description of the composition of the sample as possible.

Clarifying what identifies the subjects as elite should lead to greater transparency — and thus easier accumulation of knowledge — for the field, as well as benefits for the researcher, since the type of elite one has access to has implications for everything from the recruitment process (best practices for recruiting national security professionals might differ tremendously from ones targeted towards academics or wealthy citizens), to the design of the study (who exactly one's elites are might change what covariates ought to be measured as well as what experimental stimuli would be appropriate). Similarly, clarifying how a given set of elites match the theory will have commensurate benefits for researchers, as it will nudge in the direction of clearer specification of overall research design considerations, particularly the degree to which the empirical test fits the theory.

For example, a survey experiment in American politics testing a theory of Congressional responsiveness to public opinion using a heterogeneous sample of political elites that includes some federal-level politicians but consists mostly of local and state legislators might note that their sample includes some actors who match the theory exactly (based on their experience in Congress), but also includes respondents whose eliteness comes from their general occupations in politics. A survey experiment in IR testing a theory of presidential decision-making using a sample of former national security bureaucrats might note that the sampled elites — national security professionals — have domain-relevant expertise, were identified via their former occupation and are a relatively close "match" to the theory, with the addendum that former bureaucrats might differ from leaders elected to high office on some theoretically important dimensions, indicating what implications these differences might have for the inferences being made. Authors might similarly note that even if they were able to get former presidents to participate in their study, there are other ways in which the experimental design will invariably simplify reality (for example, the absence of time pressure, and the nature of the stakes in the scenario).

5. HOW TO UTILIZE ELITE EXPERIMENTS IN THE CONTEXT OF BROADER RESEARCH DESIGNS

Under what conditions are elite experiments likely to produce valuable inferences and advance our knowledge in the field? Two ways of thinking about this question present themselves: a more narrow conception focused only on when elite studies are particularly valu-

able, and a broader statement of how these studies might fit into larger research designs in a manner that makes the whole greater than the sum of its parts. We discuss each in turn.

5.1. When are elite studies valuable?

Two common rationales offered for elite experiments are substantive justifications — which center on whether a given study provides further evidence in favor or against a theory — and methodological justifications, which invoke concepts of external validity (the degree to which a finding extends to other settings or populations; McDermott 2011, 34). Both are reasonable in isolation — who could argue against increasing external validity of a research finding through replication on a unique sample? — but less helpful in adjudicating *relative* value of a proposed elite study against the alternative options. To that end, we propose a more systematic analysis of the value of elite studies that borrows from the classic OTUS framework (Cronbach & Shapiro 1982, see also Findley et al. 2021). Our discussion below focuses on two of those dimensions: treatments and units.

OTUS: Outcomes, treatments, units, and settings

A first dimension of studies that can help motivate the utility of elites concerns the experimental treatment itself — the "T" in OTUS. We argue that elite studies are particularly informative to the extent that they test theories that directly implicate elites' domain-specific expertise and experience. If, for example, we expect that one of the cognitive advantages of elites is they make better use of heuristics (Hafner-Burton et al. 2013), it follows that experiments will be the most useful if they mirror the tasks in which these heuristics should materialize. This is a relatively easy condition to satisfy, since most elite experiments test theories that either are directly about elites or have implications for what we ought to expect of elite behavior, judgment, preferences, etc. A smaller category of elite experiments in political science test more general theories of behavior or psychology, and these are motivated almost entirely on methodological grounds. Since there is nothing in theories of heuristic decision making that suggests elites would be differentially likely than non-elites to use heuristics, studies extending this research to elites (e.g., Stolwijk & Vis 2020) motivate their contribution exclusively via external validity considerations (in a storied tradition of extending behavioral research to unusual samples, e.g., Gigerenzer & Kurzenhaeuser 2005).

Second, elite studies are the most useful where the elites (the "units") being studied most closely resembles the target population implied by the theory. The use of elite subjects in and of itself rarely "solves" the issue of external validity — which, in any case, is better thought of a property of research programs rather than individual studies — but can be valuable to the extent that subjects correspond to the actors in the theory. On what dimensions they ought to correspond depends, of course, on the theory as well as the researcher's conception of "eliteness." For some researchers, domain-specific knowledge is critical while for others eliteness is conferred via demographic or even psychological variables.

Focusing attention on the correspondence between the actors in a given theory and one's sample is valuable on its own, but additionally clarifies the necessity of extrapolation even when using elite samples. Many elite experiments in IR, for example, are fielded on samples of legislators (Findley et al. 2017a) or mid-career military officers (Mintz et al. 2006) who, while more elite than first year college students, are still invariably removed from the high-ranking members of the executive branch who are the primary decision makers in most instances of foreign policy decision-making (Saunders 2022).

Assessing correspondence between one's elite subjects and the actors in a theory hinges

on not just identifying the elites in the theory, but making the case that the elites in one's samples are representative of that broader population. This is usually discussed as an issue of response rates — which vary by study mode, with in-person studies featuring the highest response rates (Vis & Stolwijk 2020, 3-4) — but even high response rates do not necessarily imply representativeness, and assessing this quantity is simpler with some types of elites than others. When fielding surveys or experiments on high-ranking elites from well-defined populations – e.g., current and former members of the Israeli Knesset (Renshon et al. 2020) – it is relatively straightforward to assess how representative the respondents are, due to the presence of accessible benchmarks (e.g., biographical and demographic data on legislators is often publicly available, allowing researchers to compare the observable characteristics of legislators who respond to the survey with those of the population of legislators as a whole). In contrast, this type of biographic data is less likely to be available for samples of mid-level bureaucrats, military officers or local elites. The accessibility of biographical and demographic data on more prominent political elites is also a boon to the study of biographical approaches to the study of leaders (Krcmaric et al. 2020), since researchers can augment the data they collect from elite participants with additional observational or behavioral measures they collect outside the confines of the study, allowing for shorter instruments, thereby increasing the likelihood of busy elites participating in the study.

5.2. Integrating elite studies into research designs

Related to the question of whether elite studies are worth the effort is a broader issue of how they fit into larger research designs alongside other experimental or non-experimental designs. Research designs that incorporate elites can be categorized along a number of dimensions. For projects with multiple experiments, one key feature is whether a common set of treatments is administered across both elite and non-elite samples – allowing for direct comparisons between the two. Other ways of thinking about the aggregation of different methods include whether observational data is used to motivate a finding that is then tested using an elite experiment, or if, observational data is collected after an experiment has been fielded to show similar patterns in different (non-experimental) contexts or explore mechanisms suggested by the results of an experiment. We discuss each of these approaches below.

One increasingly common method of using elite studies are in combination with other experimental studies, either on elites or mass samples. Some of these studies include *paired experiments*, where an overlapping set of treatments and outcome measures are administered on both an elite and a mass sample simultaneously, to test whether the patterns obtained on a mass sample also hold in an elite one (Kertzer 2021). For example, Teele et al. (2018) field conjoint experiments on samples of American local and state legislators, as well as the American mass public, to study gender bias in candidate preference. Others include conceptual replications, like Karpowitz et al. (2017), who combine a field experiment on Republican precinct chairs with a survey experiment on Republican primary election voters, to test how messages from party leaders affects' female candidates' electoral success. Others include *complementary experiments*, where a non-overlapping set of treatments and/or outcome measures are administered on mass and elite samples, reflecting the different roles each actor plays in politics. For example, Butler & Powell (2014) field survey experiments on voters to study how they respond to party brands, alongside survey experiments on state legislators to see how party brands affect legislators' votes.

Paired experiment: An experiment where a common set of treatments and outcome measures are administered on both an elite and mass sample

Complementary experiment: An experiment where a nonoverlapping set of treatments and outcome measures are administered on both an elite and mass sample

While targeted comparisons are useful — both for more prosaic purposes of replication as well as exploring differences between elite and mass public populations, a word of caution is in order. One common feature of paired studies is an emphasis on ways in which the elites under examination differ from the mass public. Dietrich et al. (2021b), for example, suggests that "scholars should consider investing in elite experiments precisely because research suggests that elites behave fundamentally differently from non-elites." Indeed, it is not hard to find evidence suggestive of differences between elite and non-elite samples in paired setups, such as that provided by Mintz et al. (2006) comparing military elites to an undergraduate population in a decision-making experiment. And yet, as Kertzer (2021) points out, "eliteness" itself is never itself causally identified in paired studies thus far. This means that scholars sometimes over-interpret differences between elite and mass samples, attributing effects to domain-specific expertise that may instead be a function of compositional differences (such as elite samples in many countries typically consisting of older, wealthy men). Researchers also sometimes fail to distinguish between differences in intercepts versus differences in slopes, focusing on average differences in an outcome variable between the two groups, rather than whether the effects of the treatment itself on that outcome measure actually differ between the two groups. These tendencies have arguably skewed the conclusions we have drawn from many of these studies.

Other work combine elite experiments with non-experimental data of various kinds, which can be collected either before or after the experiment has been fielded. An example of the former type of design is one in which observational survey data is used to motivate a problem by drawing attention to variation in political outcomes: for example, Flavin and Hartney's (2017) analysis of California school board elections finds correlational evidence that school board members are only punished based on achievement levels of white students in their districts. They then use this finding to motivate a list experiment, which obtains causal evidence that electoral pressures depend on the race of the students in question.

In other cases, the observational data analysis is designed to build on experimental findings after the fact, often by demonstrating that a corresponding empirical pattern exists outside of the experimental context. For example, Hemker & Rink (2017) use a conjoint experiment to show that response rates among German welfare offices to German vs. non-German requests were indistinguishable but non-Germans received lower quality responses. This pattern was substantiated using observational data that compared welfare offices run by local governments (as opposed to national bureaucracies) to show similar patterns of discrimination. Other work combines elite experiments with non-experimental sources of data, such as qualitative content analysis, to explore additional implications of their theory (Distelhorst & Hou 2017). Variants of this approach have also used clever designs in which treatments are administered to legislators, and outcomes measured using publicly available spending and parliamentary attendance data (Ofosu 2019).

6. ACCESS AND DESIGN IN ELITE EXPERIMENTS

Among the central challenges one confronts in elite experiments relates to design and recruitment, particularly since, as the quantitative literature review shows, scholars fielding elite experiments are more likely to recruit participants themselves than scholars fielding studies on the mass public – although the rise of firms like CivicPulse, who maintain panels of (usually local) political elites to which researchers can purchase access (e.g., Shaffer et al. 2020) raises the possibility that more developed infrastructures for elite studies will emerge

in the future.

6.1. Recruitment

Over the years, any number of practical guides have been written — often from the perspective of survey research or interviewing in American politics or field experiments in the study of institutions — that bear on elite recruitment and access. Aggregating across these, there are a number of areas of agreement:

1. The first contact with elite respondents is incredibly important (Goldstein 2002, Dahlberg 2007, Efrat 2015)
2. Researchers should follow up with their elite respondents multiple times (Vis & Stolwijk 2020, 11)
3. Researchers should address the issue of anonymity and duration of the study directly and from the beginning (Goldstein 2002, 670; see also Dietrich et al. 2021b) in order to increase one's sample size.
4. Researchers should carefully consider what the elites might get out of the research (Loewen et al. 2010, Dietrich et al. 2021a), such as personalized feedback on their decision-making (Carnevale et al. 2011) or a briefing (Dietrich et al. 2021b) on the overall results.

In other areas, there is less agreement, such as whether and what type of incentives are appropriate: Dietrich et al. (2021b) suggest monetary incentives, while Godwin (1979) reports that monetary incentives didn't increase the sample size of elites responding to a mail survey and besides, some elites will have rules about accepting compensation (Dietrich et al. 2021a), or find monetary incentives insulting (Renshon 2015, 674). There is also divided advice on questions like whether using "forced choice" designs will annoy elites — Maestas et al. (2003) suggests it will, dissenting from Godwin's (1979) reassurance that it will not — and whether to approach leaders of the elite group (e.g., a local chairperson or party leader) to get their stamp of approval prior to fielding the study. On the latter issue, Dietrich et al. (2021b) argue that doing so will generate "buy-in" while Vis & Stolwijk (2020, 11) highlight the risks of such an approach (if the leaders decline to cooperate, the researchers will have potentially lost a lot of respondents).

Poisoning the well: When respondents' negative experiences from one researcher's study decreases the likelihood of respondents participating in future studies by other researchers

A related issue is whether the dangers of "poisoning the well" of elite subjects might be more pronounced than among other groups. Political scientists already sit uneasily in the middle of a debate on deception in which psychologists and economists represent opposite poles. To paraphrase McDermott (2013, 605-606), psychologists tend to believe that a little deception isn't so bad and find that subjects may even enjoy participating in research that uses deception more (Sharpe et al. 1992), while economists generally prohibit deception and argue that it generates suspicion (Ortmann & Hertwig 2002) and affects selection into future studies (Jamison et al. 2008). While it is not yet clear how much danger there is in general of poisoning the well of subjects, it's easy to see that any spillover effects that would occur in the general population might be magnified if elites are more attentive or have more interconnected social networks compared to undergraduate or MTurk samples. These concerns are also broader for elites (who as prominent figures may worry about being embarrassed in some way in a manner that is dissimilar to undergraduates) and deeper. In our experience, we have encountered political elites who were aware of previous research that used deception, such that we had difficulty convincing potential elite respondents that we

would not use deception even when we pointed directly at a sign in the lab that proclaimed that it was a "no-deception lab!"

One might wish for more guidance based on systematic empirical work, and less based on anecdotes and hunches about why one approach worked and another did not. To that end, we encourage researchers to incorporate manipulations designed to assess the relative efficacy of different approaches to recruiting political elites, varying the presence of incentives, follow-up contact, and non-monetary compensation as secondary objects of study in their research. Until such time, we are left with a range of helpful anecdotes and experiences worth considering how to apply them in one's own research. Given the considerable heterogeneity of elite studies, however, researchers should avoid uncritically adopting recruitment protocols from other elite studies, and should instead solicit feedback and engage in discussions with the relevant elite population. There is almost no telling, *ex ante*, what constraints or expectations a given subpopulation might have — military officials enrolled in an executive education program at a research university have different expectations about compensation than if they were approaches at their "day job" — so the exigencies of piloting and listening to feedback are perhaps even stronger for recruiting elite samples than they are in other cases.

6.2. Sample size and representativeness

The most notable feature of elite samples also has the largest implications for design: as noted in our literature review, elite samples are typically smaller than mass samples, and more difficult to access, thereby raising their cost, and sometimes even precluding follow-up studies. One technique that scholars often use to increase their sample size in elite studies is to obtain heterogeneous samples of elites, aggregating across multiple elite populations. In American politics, for example, survey experiments are often fielded on a pooled sample of local, state, and federal legislators (Teele et al. 2018), or mixes of legislators and staffers (Malhotra et al. 2019). In International Relations, experimentalists sometimes run studies on the "foreign policy establishment" or "foreign policy opinion leaders", bringing to mind the notion of a "Blob" in foreign policy stretching from Congress to think tanks to business leaders to the ivory tower (Busby et al. 2020a, Kertzer et al. 2021a). In general, these studies don't find much evidence of heterogeneous treatment effects across elite types (though see Gift & Monten 2021), although many of these subsets are often relatively small.

As always, tradeoffs abound: despite the advantages inherent in pooling across different types of elites and the lack of evidence that doing so results in heterogeneous treatment effects within a study, heterogeneous elites samples — e.g., mixes of civil servants, former legislators and staff, and corporate executives (Hafner-Burton et al. 2014) — make it more difficult to assess representativeness. In a sample of heterogeneous elites, should researchers weight each group of elites equally, or according to other population benchmarks? If researchers choose the latter in an effort to estimate the population average treatment effect (PATE), caution is in order, but methods are available to estimate conservative bounds (Miratrix et al. 2018) and describe the choices in a transparent manner (Franco et al. 2017). As is the case with all methods then, elite experiments and surveys face a number of tradeoffs. Higher-ranking officials may be less likely to participate in a study, but researchers are more likely to be able to gather observational or behavioral data about the elites outside of the study to incorporate into the analysis than is the case with lower-ranking elites. Lower-ranking officials may have more time for researchers, but are also less likely to have

this type of data available. Heterogeneous samples of elites allow researchers to bolster their statistical power due to obtaining larger sample sizes, but also make it harder for the researcher to assess the representativeness of the sample to the target population.

The nature of elite samples — often smaller and more costly to access — imply a number of best practices for our designs. For example, smaller samples suggest the wisdom of more extensive piloting to hone the efficiency and precision of treatments in advance of fielding, and the use of pilot data to generate realistic estimates of statistical power. The latter consideration is often advised, yet rarely followed (Sedlmeier & Gigerenzer 1989). New tools such as DeclareDesign (Blair et al. 2019) are a useful tool in this vein, but its output will only be as good as the preparation that goes into it: there is no substitute for piloting to help choose and revise treatments, as well as establish clear contrasts across arms of a study. Given the danger of "burning" an already small sample by piloting on elites, we recommend using easier-to-access samples for this purpose — still useful in testing and honing design choices and treatments — alongside informal consultation with a small number of elites to discuss sample-specific issues.

Of course, some designs will be more efficient than others in the context of smaller-than-preferred samples. Yarhi-Milo et al. (2018) field a within-subjects design on their sample of legislators from the Israeli Knesset, out of a concern that a between subjects design would require a much larger sample size than would be practical; in the context of economics experiments, Bellemare et al. (2014) estimate that between-subjects designs require between 4-8 times as many subjects as within-subjects design to reach an equivalent level of statistical power. The typical concern with such a strategy is that "consistency pressures" will lead to reduced treatment effects, though Clifford et al. (2021) show that both within- and between-subject versions of "repeated measures designs" yield significant benefits in precision and statistical power over more traditional designs that measure outcomes only post-treatment. Conjoint experiments are a common method of leveraging both between- and within-subject designs, and recent advances suggest that earlier concerns about satisficing with large numbers of tasks (Bansak et al. 2018) may have been misplaced (though these designs do make it more difficult to analyze subgroup preferences than more traditional designs, Leeper et al. 2020). Nonetheless, many elite samples may not be willing to allocate the kind of time to the researcher necessary to administer conjoint experiments with larger numbers of choice tasks.

Just as a fear of inducing consistency bias may have been overblown, other research on the meta-effects of different aspects of design suggest that elite experimenters need not be overly cautious. Mummolo & Peterson (2019) find little evidence for demand effects in survey experimental designs, freeing up researchers to take the advice of Grose (2021, 157) to use "meaningful, fairly infrequent and bold interventions" without fear. Another typical concern is that certain subjects might be particularly sensitive for elite respondents even with assurances of anonymity and non-identifiable data. In those cases, researchers' instinct is often to design their treatments and vignettes in a manner that won't provoke unease or increased attrition by making vignettes hypothetical or replacing the names of countries or actors with fictional types. While some worry that hypothetical questions will generate hypothetical answers, Brutger et al. (2020) show that there are fewer tradeoffs associated with abstraction in experimental design than is commonly believed.

6.3. How to study elites without access to elites

A final consideration in the study of elites is how researchers might operate in a world of scarcity, where access to high-ranking political elites is either costly, infeasible, or raises ethical concerns (Nathan & White 2021). Here, two approaches suggest themselves as potential avenues for researchers who may not have access to elites. These two approaches address recruitment and design, respectively.

The first set of options for researchers occurs at the recruitment stage. Here, there are several options depending on one's level of access and resources. Given some level of access to elites, the first option is to combine smaller samples rather than hold out for access to one large, uniform group of elite respondents (e.g., Teele et al. 2018, Kertzer et al. 2021a). As noted above, there has thus far been little evidence of heterogenous treatment effects among studies that use pooled samples of elites (at varying levels of eliteness), so aggregating across multiple (smaller) elite samples is one possible route. Another is to broaden the definition of elites outside of what is directly specified by one's theory: IR scholars might, for example, consider the use of retired military officers, individuals enrolled in ROTC programs, or officers in professional military education programs (e.g., Jost et al. 2017, Friedman et al. 2017); comparativists might expand their aperture to look at both former and current legislators, rather than just individuals currently in elected office – although care should be taken if there have been substantial partisan shifts between legislative terms. Or, if one's model of eliteness is closer to compositional models that focus on socioeconomic characteristics of elites, older, more highly educated, wealthier respondents might function as appropriate stand-ins for elites. The specific demographic characteristics to subset on will depend on the particular elite population under investigation.

A related approach that bridges the gap between recruitment and design is to create one's own elites within the context of the study. Here, we recommend that scholars consider the theoretically-relevant dimensions along which elites may differ from non-elites and build those into studies whose subjects are not traditional elites. For example, some cognitive models of elites suggest that one critical aspect of eliteness is domain-specific experience, in which case researchers can train respondents in ways that create home-grown experts in the lab, such as Tingley (2011) who compared the behavior of students who received a lecture on repeated games to that of naive subjects. Other models of eliteness suggest that there are aspects of elite psychology that differ systematically from non-elites, whether those qualities are traits like patience (Hafner-Burton et al. 2014), greater sense of power (Renshon 2015) or other factors. Just as one might train respondents and give them the domain-specific knowledge they need to be more expert, one might either incentivize or prime dispositional traits, ideally as part of assignment into an arm of a randomized study in order to properly identify the differences between more traditional subjects and those assigned to be elites. The virtue of these approaches is that they force theoretical clarity onto the researcher about what makes elites special – and render explicit the tradeoffs with elite experiments that are often left unstated. They also provide more causal leverage in identifying the effect of traits, knowledge, or experience that we think make actors elite in the first place.

A final option involves the design of the experiment, and addresses not the units — elites — themselves but asks researchers to consider whether the most important feature to be tested experimentally involve the actors or rather, the setting or context in which they make decisions. If the former is critical, then the options above are available to researchers. If the latter is important, it's possible to consider how one might build features from the

context implicated by the theory into the experimental design. For example, attributes of the context like time pressure, group processes, accountability and uncertainty are all commonly-invoked elements of elite decision-making and might easily be incorporated into experimental designs. Modeling these group dynamics is particularly valuable given the extent to which much elite decision-making often takes place in small group rather than individual contexts (Saunders 2017, Kertzer et al. 2021b).

7. CONCLUSION

In a foundational text in experimental economics, Roth (1986) noted that laboratory experiments traditionally had one of three different purposes: testing the propositions of formal theories, demonstrating empirical regularities, or "whispering in the ears of princes." In the social sciences in recent years there has since been a shift, as experimenters have sought not only to supplicate royal earlobes, but study them directly, fielding experiments in which political elites are the respondents themselves. An acceleration in the number of elite studies, however, has outpaced our collective understanding of best practices, both in terms of the logistics of administering elite experiments, but also how to use elite studies to best advance knowledge in the discipline.

In this article, we made a number of recommendations. First, scholars studying political elites should be explicit about who the relevant political elites are for the questions they are studying, justifying why the elite sample they use is well suited to testing their theory. Elite studies are ultimately the most useful where the elites being studied most closely resemble the target population implied by our theoretical frameworks. At the same time, however, there will invariably be disjunctures between the kinds of elites that researchers are most likely able to have access and the ultimate high-level decision-makers studied by many of our theories. Being clear about these disjunctures when they arise will help bolster the credibility of the findings.

Second, elite experiments are particularly valuable when they are designed to test theories that directly implicate elites' domain-specific expertise and experience. In this sense, the treatments respondents are being presented with are as important as the question of who the respondents themselves are. Recruiting a sample of high-ranking political elites and presenting with treatments or scenarios unrelated to the tasks they carry out in political life will lead to less helpful inferences.

Third, attention should be paid to the role of elites in broader research designs. Researchers can replicate their studies on elites with studies on mass samples, as in paired experiments, or in complementary experiments where researchers test different microfoundations of a broader theoretical framework by fielding different experiments on each sample. Researchers can similarly combine elite experiments with non-experimental data, either to motivate an experimental design or to validate its findings.

Fourth, given the number of contradictory anecdotes and intuitions about the best way to recruit elite respondents, researchers should consider studying questions of elite recruitment experimentally, which will provide greater insight about how different recruitment approaches, incentive schemes, follow-up contact, and non-monetary forms of compensation actually affect the quantity and quality of elite participation.

Fifth, as with all experiments, researchers fielding elite experiments face a number of tradeoffs between sample size, eliteness, and representativeness: e.g., heterogeneous elite samples can typically be larger in size, but also more likely to have ill-defined popula-

tion frames; the more proximate respondents are to positions of political power, the less time they'll have for the study, and the more pressing concerns about statistical power will become. While some experimental designs, like those that incorporate within-subjects components, are better suited to smaller samples, tradeoffs remain inevitable, such that researchers should be explicit about how they navigated these tradeoffs when designing their study.

Finally, given challenges of accessing elite samples, there are a number of alternative research designs scholars without access to elite samples of sufficient size may wish to consider, from designs that pool heterogeneous groups of political elites, to designs that induce domain-specific expertise among a random subset of respondents. Broadening the elite experimental literature in this manner may help increase the rate of accumulation of knowledge.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Hohyun Yoon for research assistance, and David Broockman, Guy Grossman, Brendan Nyhan, Elizabeth Saunders, Dustin Tingley, and Jessica Weeks for helpful feedback. Authors listed in alphabetical order.

LITERATURE CITED

Arceneaux K, Dunaway J, Soroka S. 2018. Elites are people, too: The effects of threat sensitivity on policymakers' spending priorities. *PloS one* 13(4)

Auerbach AM, Thachil T. 2020. Cultivating clients: Reputation, responsiveness, and ethnic indifference in india's slums. *American Journal of Political Science* 64(3):471–487

Axelrod R. 1976. Structure of decision: The cognitive maps of political elites. Princeton university press

Baekgaard M, Christensen J, Dahlmann CM, Mathiasen A, Petersen NBG. 2019. The role of evidence in politics: Motivated reasoning and persuasion among politicians. *British Journal of Political Science* 49(3):1117–1140

Bahador B, Entman R, Knüpfer C. 2019. Who's elite and how the answer matters to politics. *Political Communication* 36(1):195–202

Bansak K, Hainmueller J, Hopkins DJ, Yamamoto T. 2018. The number of choice tasks and survey satisficing in conjoint experiments. *Political Analysis* 26(1):112–119

Banuri S, Dercon S, Gauri V. 2019. Biased Policy Professionals. *The World Bank Economic Review* 33(2):310–327

Bayram AB. 2017. Due deference: Cosmopolitan social identity and the psychology of legal obligation in international politics. *International Organization* 71(S1):S137–S163

Bellemare C, Bissonnette L, Kröger S. 2014. Statistical power of within and between-subjects designs in economic experiments. CRREP working paper 2014-03

Blair G, Cooper J, Coppock A, Humphreys M. 2019. Declaring and diagnosing research designs. *American Political Science Review* 113(3):838–859

Broockman DE, Ferenstein G, Malhotra N. 2019. Predispositions and the political behavior of

american economic elites: Evidence from technology entrepreneurs. *American Journal of Political Science* 63(1):212–233

Broockman DE, Skovron C. 2018. Bias in perceptions of public opinion among political elites. *American Political Science Review* 112(3):542–563

Brutger R, Kertzer JD, Renshon J, Tingley D, Weiss CM. 2020. Abstraction and detail in experimental design. Working Paper.

Busby J, Kafura C, Monten J, Tama J. 2020a. Multilateralism and the use of force: Experimental evidence on the views of foreign policy elites. *Foreign Policy Analysis* 16(1):118–129

Busby J, Kafura C, Smeltz D, Tama J, Monten J, et al. 2020b. Coming together or coming apart? attitudes of foreign policy opinion leaders and the public in the trump era. Chicago Council on Global Affairs Policy Brief

Bussell J. 2020. Shadowing as a tool for studying political elites. *Political Analysis* 28(4):469–486

Butler DM, Crabtree C. 2021. Audit studies in political science. *Advances in Experimental Political Science* :42

Butler DM, Kousser T. 2015. How do public goods providers play public goods games? *Legislative Studies Quarterly* 40(2):211–240

Butler DM, Nickerson DW. 2011. Can learning constituency opinion affect how legislators vote? results from a field experiment. *Quarterly Journal of Political Science* 6(1):55–83

Butler DM, Powell EN. 2014. Understanding the party brand: Experimental evidence on the role of valence. *Journal of Politics* 76(2):492–505

Carey JM, Helmke G, Nyhan B, Sanders M, Stokes S. 2019. Searching for bright lines in the trump presidency. *Perspectives on Politics* 17(3):699–718

Carnes N, Lupu N. 2015. Rethinking the comparative perspective on class and representation: Evidence from latin america. *American Journal of Political Science* 59(1):1–18

Carnevale JJ, Inbar Y, Lerner JS. 2011. Individual differences in need for cognition and decision-making competence among leaders. *Personality and Individual Differences* 51(3):274–278

Clifford S, Sheagley G, Piston S. 2021. Increasing precision without altering treatment effects: Repeated measures designs in survey experiments. *American Political Science Review* :1–18

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

Cronbach LJ, Shapiro K. 1982. Designing evaluations of educational and social programs. Jossey-Bass,

Dahlberg S. 2007. Web-based expert surveys. QoG Working Paper Series

Dal Bó E, Finan F, Folke O, Persson T, Rickne J. 2017. Who becomes a politician? *Quarterly Journal of Economics* 132(4):1877–1914

Dietrich S, Hardt H, Swedlund HJ. 2021a. Elite experiments: Strengthening scholarship while bridging the gap. Duck of Minerva

Dietrich S, Hardt H, Swedlund HJ. 2021b. How to make elite experiments work in international relations. *European Journal of International Relations* :1354066120987891

Distelhorst G, Hou Y. 2017. Constituency service under nondemocratic rule: Evidence from china. *The Journal of Politics* 79(3):1024–1040

Dynes AM, Hassell HJG, Miles MR, Preese JR. 2021. Personality and gendered selection processes in the political pipeline. *Politics & Gender* 17(1):53–73

Efrat A. 2015. Cross-national interviewing at international conferences: how to make the most of a unique research opportunity. *International Studies Perspectives* 16(3):302–311

Fatas E, Neugebauer T, Tamborero P. 2007. How politicians make decisions: A political choice experiment. *Journal of Economics* 92(2):167–196

Fenno RF. 1978. Home style: House members in their districts. Boston: Little, Brown

Findley MG, Harris AS, Milner HV, Nielson DL. 2017a. Who controls foreign aid? elite versus public perceptions of donor influence in aid-dependent Uganda. *International Organization* 71(4):633–663

Findley MG, Kikuta K, Denly M. 2021. External validity. *Annual Review of Political Science* 24(1):365–393

Findley MG, Laney B, Nielson DL, Sharman JC. 2017b. External validity in parallel global field and survey experiments on anonymous incorporation. *The Journal of Politics* 79(3):856–872

Flavin P, Hartney MT. 2017. Racial inequality in democratic accountability: Evidence from retrospective voting in local elections. *American Journal of Political Science* 61(3):684–697

Franco A, Malhotra N, Simonovits G, Zigerell L. 2017. Developing standards for post-stratification weighting in population-based survey experiments. *Journal of Experimental Political Science* 4(2):161–172

Friedman JA, Lerner JS, Zeckhauser R. 2017. Behavioral consequences of probabilistic precision: Experimental evidence from national security professionals. *International Organization* 71(4):803–826

Fuhrmann M, Horowitz MC. 2015. When leaders matter: Rebel experience and nuclear proliferation. *Journal of Politics* 77(1):72–87

George AL. 1969. The" operational code": A neglected approach to the study of political leaders and decision-making. *International studies quarterly* 13(2):190–222

George AL, George JL. 1964. Woodrow Wilson and Colonel House: a personality study. Dover Press

Gerring J, Oncel E, Morrison K, Pemstein D. 2019. Who rules the world? a portrait of the global leadership class. *Perspectives on Politics* 17(4):1079–1097

Gift T, Monten J. 2021. Who's out of touch? media misperception of public opinion on us foreign policy. *Foreign Policy Analysis* 17(1):128–139

Gigerenzer G, Kurzenhaeuser S. 2005. Fast and frugal heuristics in medical decision making. *Science and medicine in dialogue: Thinking through particulars and universals* :3–15

Godwin RK. 1979. The consequences of large monetary incentives in mail surveys of elites. *Public Opinion Quarterly* 43(3):378–387

Goemans HE, Gleditsch KS, Chiozza G. 2009. Introducing Archigos: A dataset of political leaders. *Journal of Peace research* 46(2):269–283

Goldstein K. 2002. Getting in the door: Sampling and completing elite interviews. *PS: Political Science and Politics* 35(4):669–672

Grose CR. 2014. Field experimental work on political institutions. *Annual Review of Political Science* 17:355–370–61

Grose CR. 2021. Experiments, political elites and political institutions. In *Advances in Experimental Political Science*, eds. J Druckman, DP Green. Cambridge University Press

Grossman G, Michelitch K. 2018. Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda. *American Political Science Review* 112(2):280–301

Gulzar S. 2021. Who enters politics and why? *Annual Review of Political Science* 24(1):253–275

Hafner-Burton EM, Hughes DA, Victor DG. 2013. The cognitive revolution and the political psychology of elite decision making. *Perspectives on Politics* 11(2):368–386

Hafner-Burton EM, LeVeck BL, Victor DG, Fowler JH. 2014. Decision maker preferences for international legal cooperation. *International Organization* 68(4):845–876

Hafner-Burton EM, Victor DG, LeVeck BL. 2016. How activists perceive the utility of international law. *The Journal of Politics* 78(1):167–180

Hemker J, Rink A. 2017. Multiple dimensions of bureaucratic discrimination: Evidence from german welfare offices. *American Journal of Political Science* 61(4):786–803

Hertel-Fernandez A, Mildenberger M, Stokes LC. 2019. Legislative staff and representation in congress. *American Political Science Review* 113(1):1–18

Heß M, Von Scheve C, Schupp J, Wagner A, Wagner GG. 2018. Are political representatives more risk-loving than the electorate? Evidence from German federal and state parliaments. *Palgrave Communications* 4(1):1–7

Hoffmann-Lange U. 2007. Methods of elite research. In *The Oxford handbook of political behavior*

Hoffmann-Lange U. 2008. Studying elite vs mass opinion. *The SAGE handbook of public opinion research* :53–63

Humayun F. 2021. The punisher's dilemma: Domestic opposition and foreign policy losses. Working paper

Jamison J, Karlan D, Schechter L. 2008. To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization* 68(3-4):477–488

Jost T, Meshkin K, Schub R. 2017. Socialized hawks? how selection explains military attitudes on the use of force. Working paper

Karpowitz CF, Monson JQ, Preese JR. 2017. How to elect more women: Gender and candidate success in a field experiment. *American Journal of Political Science* 61(4):927–943

Keller FB. 2016. Moving beyond factions: using social network analysis to uncover patronage networks among chinese elites. *Journal of East Asian Studies* 16(1):17–41

Kennedy C, Hartig H. 2019. Response rates in telephone surveys have resumed their decline. Pew Research Center

Kertzer JD. 2021. Reassessing elite-public gaps in political behavior. *American Journal of Political Science* Forthcoming

Kertzer JD, Busby J, Monten J, Tama J, Kafura C. 2021a. Elite misperceptions and the domestic politics of conflict. Working paper

Kertzer JD, Holmes M, LeVeck BL, Wayne CN. 2021b. Hawkish biases and group decision-making. Working paper

Kertzer JD, Renshon J, Yarhi-Milo K. 2021c. How do observers assess resolve? *British Journal of Political Science* 51(1):308–330

Kertzer JD, Tingley D. 2018. Political psychology in international relations: Beyond the paradigms. *Annual Review of Political Science* 21:319–339

Krcmaric D, Nelson SC, Roberts A. 2020. Studying leaders and elites: The personal biography approach. *Annual Review of Political Science* 23(1):133–151

Lasswell HD. 1952. The elite concept. In *The Comparative Study of Elites: An Introduction and Bibliography*. Stanford, CA: Stanford University Press, 6–21

Leeper TJ, Hobolt SB, Tilley J. 2020. Measuring subgroup preferences in conjoint experiments. *Political Analysis* 28(2):207–221

Lin-Greenberg E. 2021. Soldiers, pollsters, and international crises: Public opinion and the military's advice on the use of force. *Foreign Policy Analysis* Forthcoming

Lin-Greenberg E, Pauly R, Schneider J. 2020. Wargaming for political science research. Working paper

Linde J, Vis B. 2017. Do politicians take risks like the rest of us? an experimental test of prospect theory under MPs. *Political Psychology* 38(1):101–117

List JA, Mason CF. 2011. Are ceos expected utility maximizers? *Journal of Econometrics* 162(1):114–123

Loewen PJ, Rubenson D, 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

Lupia A, McCubbins MD. 1998. The democratic dilemma: Can citizens learn what they need to know? New York: Cambridge University Press

Maestas C, Neeley GW, Richardson Jr LE. 2003. The state of surveying legislators: Dilemmas and suggestions. *State Politics & Policy Quarterly* 3(1):90–108

Mahdavi P, et al. 2017. Scraping public co-occurrences for statistical network analysis of political elites. *Political Science Research and Methods* 7:385–392

Malesky E, Schuler P, Tran A. 2012. The adverse effects of sunshine: a field experiment on legislative transparency in an authoritarian assembly. *American Political Science Review* 106(4):762–786

Malhotra N, Monin B, Tomz M. 2019. Does private regulation preempt public regulation? *American*

Political Science Review 113(1):19–37

McDermott R. 2011. Internal and external validity. *Cambridge handbook of experimental political science* :27–40

McDermott R. 2013. The ten commandments of experiments. *PS: Political Science & Politics* 46(3):605–610

Meisel JH. 1962. The myth of the ruling class: Gaetano Mosca and the "elite". Ann Arbor, MI: University of Michigan

Miller WE, Stokes DE. 1963. Constituency influence in congress. *The American Political Science Review* 57(1):45–56

Mintz A, Redd SB, 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

Miratrix LW, Sekhon JS, Theodoridis AG, Campos LF. 2018. Worth weighting? how to think about and use weights in survey experiments. *Political Analysis* 26(3):275–291

Moyser G, Wagstaffe M. 1987. Studying elites: theoretical and methodological issues. In *Research Methods for Elite Studies*, eds. G Moyser, M Wagstaffe. Allen & Unwin, 1–24

Mummolo J, Peterson E. 2019. Demand effects in survey experiments: An empirical assessment. *American Political Science Review* 113(2):517–529

Nair D. 2021. "hanging out" while studying "up": Doing ethnographic fieldwork in international relations. *International Studies Review* Forthcoming

Nathan NL, White A. 2021. Experiments on and with street-level bureaucrats. In *Handbook of Advances in Experimental Political Science*, eds. J Druckman, DP Green. Cambridge University Press

Neumann IB. 2012. At home with the diplomats: Inside a european foreign ministry. Cornell University Press

Nyhan B, Reifler J. 2015. The effect of fact-checking on elites: A field experiment on us state legislators. *American Journal of Political Science* 59(3):628–640

Ofosu G. 2019. Do fairer elections increase the responsiveness of politicians? *American Political Science Review* 113(4):963–979

Oldendick RW, Bardes BA. 1982. Mass and elite foreign policy opinions. *Public Opinion Quarterly* 46(3):368–382

Ortmann A, Hertwig R. 2002. The costs of deception: Evidence from psychology. *Experimental Economics* 5(2):111–131

Page BI, Bartels LM, Seawright J. 2013. Democracy and the policy preferences of wealthy americans. *Perspectives on Politics* 11(1):51–73

Palacios-Huerta I, Volij O. 2009. Field centipedes. *American Economic Review* 99(4):1619–1635

Pereira MM. 2021. Understanding and reducing biases in elite beliefs about the electorate. *American Political Science Review* Forthcoming

Poole KT, Rosenthal H. 1997. Congress: A political-economic history of roll call voting. Oxford University Press on Demand

Potters J, van Winden F. 2000. Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization* 43(4):499–522

Putnam R, et al. 1976. The comparative study of political elites. Prentice-Hall

Ramey AJ, Klingler JD, Hollibaugh GE. 2016. Measuring elite personality using speech. *Political Science Research and Methods* 7(1):1–22

Renshon J. 2015. Losing face and sinking costs: Experimental evidence on the judgment of political and military leaders. *International Organization* 69(3):659–695

Renshon J, Yarhi-Milo K, Kertzer JD. 2020. Democratic reputations in crises and war. Working paper.

Ricart-Huguet J. 2019. The historical turn in the comparative study of political elites. *Newsletter*

of the Organized Section in Comparative Politics of the American Political Science Association 29(2):86–93

Rivera SW, Kozyreva PM, Sarovskii EG. 2002. Interviewing political elites: lessons from russia. *PS: Political Science and Politics* 35(4):683–688

Rodríguez-Teruel J, Daloz JP. 2018. Surveying and observing political elites. In *The Palgrave Handbook of Political Elites*. Springer, 93–113

Roth AE. 1986. Laboratory experimentation in economics. *Economics and Philosophy* 2:245–273

Saunders EN. 2011. Leaders at war: How presidents shape military interventions. Ithaca, NY: Cornell University Press

Saunders EN. 2017. No substitute for experience: presidents, advisers, and information in group decision making. *International Organization* :S219–S247

Saunders EN. 2022. Elites in the making and breaking of foreign policy. *Annual Review of Political Science* 25

Sedlmeier P, Gigerenzer G. 1989. Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin* 105(2):309–316

Shaffer R, Pinson LE, Chu JA, Simmons BA. 2020. Local elected officials' receptivity to refugee resettlement in the united states. *Proceedings of the National Academy of Sciences* 117(50):31722–31728

Sharpe D, G. Adair J, Roese NJ. 1992. Twenty years of deception research: A decline in subjects' trust? *Personality and Social Psychology Bulletin* 18(5):585–590

Sheffer L, Loewen P. 2019. Electoral confidence, overconfidence, and risky behavior: Evidence from a study with elected politicians. *Political Behavior* 41(1):31–51

Sheffer L, Loewen PJ, Soroka S, Walgrave S, Sheafer T. 2018. Nonrepresentative representatives: an experimental study of the decision making of elected politicians. *American Political Science Review* 112(2):302–321

Sherman GD, Lerner JS, Josephs RA, Renshon J, Gross JJ. 2016. The interaction of testosterone and cortisol is associated with attained status in male executives. *Journal of personality and social psychology* 110(6):921

Stolwijk S, Vis B. 2020. Politicians, the representativeness heuristic and decision-making biases. *Political Behavior* :1–22

Teele DL, Kalla J, Rosenbluth F. 2018. The ties that double bind: Social roles and women's underrepresentation in politics. *American Political Science Review* 112(3):525–541

Teigen M, Karlsen R. 2019. Influencing elite opinion on gender equality through framing: a survey experiment of elite support for corporate board gender quotas. *Politics & Gender* 16(3):792 – 815

Tetlock PE, Lebow RN. 2001. Poking counterfactual holes in covering laws: Cognitive styles and historical reasoning. *American Political Science Review* 95(4):829–843

Tingley DH. 2011. The dark side of the future: An experimental test of commitment problems in bargaining. *International Studies Quarterly* 55(2):521–544

Tomz M, Weeks J, Yarhi-Milo K. 2020. Public opinion and decisions about military force in democracies. *International Organization* 74(1):119–143

van der Meij L, Schaveling J, van Vugt M. 2016. Basal testosterone, leadership and dominance: A field study and meta-analysis. *Psychoneuroendocrinology* 72:72–79

Vis B, Stolwijk S. 2020. Conducting quantitative studies with the participation of political elites: best practices for designing the study and soliciting the participation of political elites. *Quality & Quantity* :1–37

Walgrave S, Sevenans J, Van Camp K, Loewen PJ. Forthcoming. What draws politicians' attention? an experimental study of issue framing and its effect on individual political elites. *Political Behavior*

Yarhi-Milo K. 2018. Who fights for reputation in international politics? leaders, resolve, and the use of force. Princeton, NJ: Princeton University Press

Yarhi-Milo K, Kertzer JD, Renshon J. 2018. Tying hands, sinking costs, and leader attributes. *Journal of Conflict Resolution* 62(10):2150–2179

Zaller JR. 1992. The nature and origins of mass public opinion. Cambridge: Cambridge University Press

Zuckerman H. 1972. Interviewing an ultra-elite. *Public Opinion Quarterly* 36(2):159–175