

*Annual Review of Political Science*

# Experiments and Surveys on Political Elites

Joshua D. Kertzer<sup>1,\*</sup> and Jonathan Renshon<sup>2,\*</sup>

<sup>1</sup>Department of Government, Harvard University, Cambridge, Massachusetts, USA;  
email: jkertzer@gov.harvard.edu

<sup>2</sup>Department of Political Science, University of Wisconsin–Madison, Madison, Wisconsin, USA;  
email: renschon@wisc.edu

ANNUAL  
REVIEWS **CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Political Sci. 2022. 25:529–50

First published as a Review in Advance on  
February 3, 2022

The *Annual Review of Political Science* is online at  
[polisci.annualreviews.org](http://polisci.annualreviews.org)

<https://doi.org/10.1146/annurev-polisci-051120-013649>

Copyright © 2022 by Annual Reviews. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information

\*These authors contributed equally to this article

## 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, the increase 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 recommendations on questions of access, recruitment, and representativeness, as well as designs that researchers can use to study “eliteness” without access to elites.

---

**Elite experiments:**

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

---

## 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, and in all of its empirical subfields 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 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 1976, Oldendick & Bardes 1982), recent shifts toward micro foundations, causal identification, survey and field experiments, and interest in leaders more generally have greatly increased the use of elite subjects in political science studies. Americanists have turned to samples of policy makers 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 (Linde & Vis 2017, Heß et al. 2018, Sheffer & Loewen 2019), and issue framing (Walgrave et al. 2018) in developed and developing countries alike. In IR, scholars have used elite samples to examine topics such as perceptions of costly signals (Yarhi-Milo et al. 2018), obligation to international law (Bayram 2017), and attitudes toward the use of force (Tomz et al. 2020). A similar surge has taken place outside of political science, most notably in research 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 various types of elite studies, including field experiments on political institutions (Grose 2014), audit studies (Costa 2017), and the use of 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 as studies where a sample of political elites has been randomly assigned by the experimenter to treatment conditions. As such, we bracket a vibrant literature utilizing elite surveys for largely descriptive purposes—for example, analyzing the policy preferences of the top 1% (Page et al. 2013), political scientists' perceptions of democratic backsliding (Carey et al. 2019), or simulations like wargames that do not contain experimental manipulations (Lin-Greenberg et al. 2022)—to focus on experimental studies where researchers use 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; thus, we set 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.

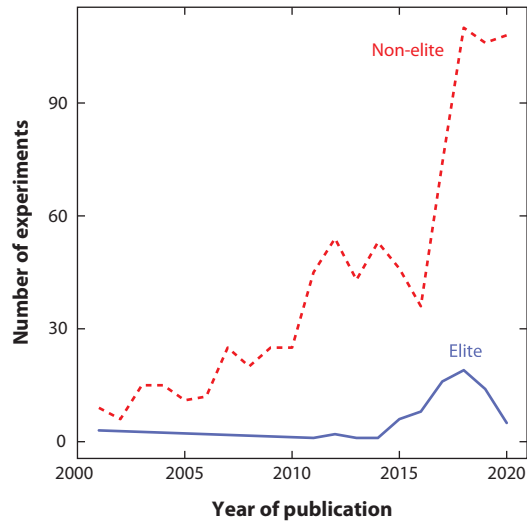
We begin by offering several observations based on a quantitative literature search of elite experiments in the “big 3” journals in political science over the last 20 years, fleshing out those developments with a broader discussion of some of the challenges that have arisen in this

literature thus far. We suggest that elite studies are most useful when they test theories directly related to elites' domain-specific expertise and experience, but that the use of elite samples in and of itself 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 to help scholars 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 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 use each sample to test a different component of the theory. Finally, we offer a practical discussion of how to study elites, focusing on how to recruit elite respondents, how to think about questions of representativeness, and how to experimentally study elites without 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.

## THE RISE OF ELITE EXPERIMENTS

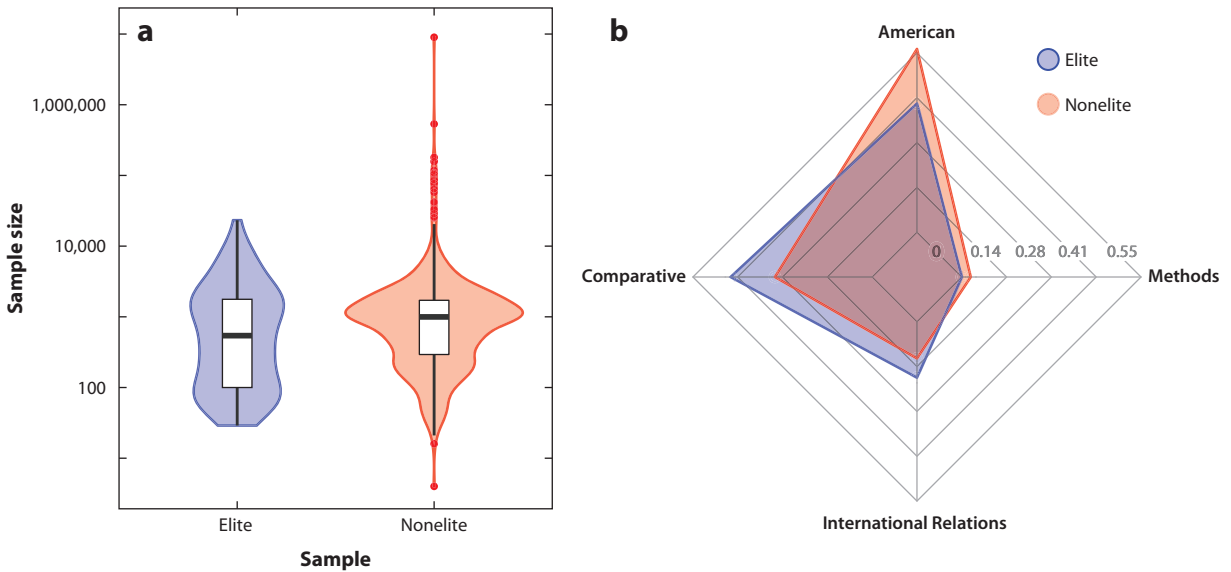
For a long time, elites and masses in political science were studied using very different methods (Kertzer & Tingley 2018). The rise of opinion polls in the 1930s (Moyser & Wagstaffe 1987, p. 5), coupled with the rise of institutionalist approaches (Ricart-Huguet 2019), caused interest in the study of political elites to wane, precisely because of an assumption that survey methods could be more fruitfully applied to mass political behavior—and despite arguments that some elites were more accessible than others [e.g., state, rather than national 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 semistructured interviews (Zuckerman 1972). This has continued today, as scholars continue to productively use a range of 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 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 *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics* from 2000 through 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 the number of published experiments on nonelite samples [nationally representative samples of mass publics, convenience samples recruited through Amazon Mechanical Turk (MTurk), 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 nonelite samples in various ways: Elite experiments tend to feature significantly smaller sample sizes, for example, and are less disproportionately focused on American politics. **Figure 3a** shows that elite experiments tend to feature a variety of



**Figure 1**

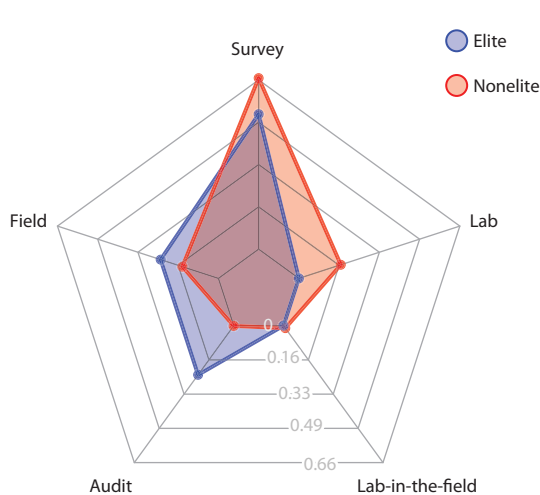
An analysis of experiments published in the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics* from 2000 through 2020 illustrates an increased number of elite experiments starting in 2014. Our unit of analysis here is the experiment (rather than the article).



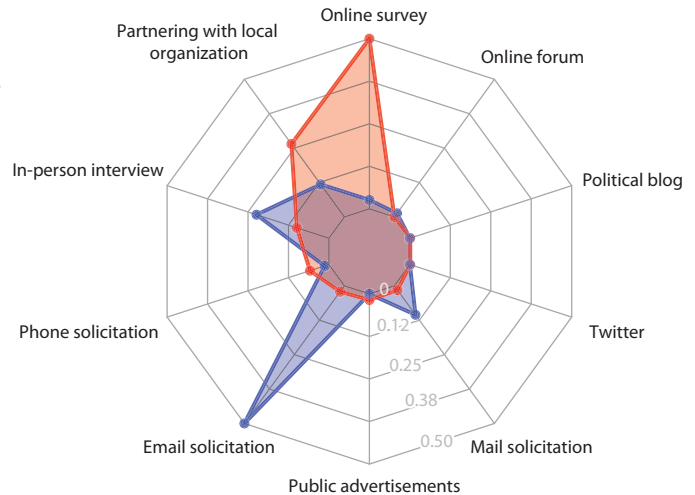
**Figure 2**

Differences between elite and nonelite experiments published in the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics* from 2000 through 2020 in terms of sample size and subfield. Panel *a* presents box plots overlaid on violin plots of the density distributions, showing that elite experiments (*blue*) tend to have smaller sample sizes than nonelite experiments (*red*) do. Panel *b* presents the proportion of elite experiments (*blue*) and nonelite experiments (*red*) published in different subfields of political science, showing that elite experiments are less dominated by American politics than nonelite experiments are.

### a Type of experiment



### b Recruitment method



**Figure 3**

Differences between elite and nonelite experiments published in the *American Journal of Political Science*, *American Political Science Review*, and *Journal of Politics* from 2000 through 2020 in terms of type of experiment and recruitment method. Panel *a* presents the proportion of elite experiments (*blue*) and nonelite experiments (*red*) by experiment type, showing that elite experiments tend to be less dominated by survey and lab experiments than nonelite experiments are. Panel *b* presents the proportion of elite experiments (*blue*) and nonelite experiments (*red*) by recruitment method, showing that respondents in elite experiments are more likely to be recruited directly by the researcher than respondents in nonelite experiments.

experimental approaches: Although a plurality of both elite and nonelite experiments are survey experiments, field and audit experiments are more common in elite than in nonelite experiments (unsurprising, given what a nonelite 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).

In line with expectations, 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 nongovernmental organizations (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 important differences as well in how the studies were fielded. One difference concerns compensation: whereas 93% of the experiments on mass samples reported providing compensation to participants (or were conducted on platforms where compensation can be inferred), only 3% of the experiments on elite samples mentioned compensating participants for their time. Another difference is in recruitment. As **Figure 3b** 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—particularly field experiments on elites—experimenters also often partner with local organizations to recruit respondents.

---

**Paired experiments:**

Experiments where a common set of treatments and outcome measures are administered on both an elite and mass sample

---

While most of the articles we surveyed included studies fielded only on either mass or elite samples, 20% of the articles that included experiments on elite samples also included experiments on mass samples, often to test how the two groups differ from one another (Kertzer 2022). 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 test theories related to the specific elites being studied; for instance, Broockman et al. (2019) use a sample of technology entrepreneurs to show that they are more conservative on issues of government regulation than their other political attitudes might predict. Other experimenters use elite samples to study more general questions about elite cognition, such as Sheffer et al. (2018), who test whether common biases identified in the judgment and decision-making literature appear 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. In our sample, 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 the mass experiments report a response rate. 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 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 it is generally lower in surveys of federal elites than municipal elites).

## CHALLENGES IN THE STUDY OF POLITICAL ELITES

Elite experiments in political science have yielded 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 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 advance knowledge in the field. This neglect has led to three sets of issues regarding how to most efficiently design and learn from elite experiments.

First, who counts as a political elite? Political scientists have used the term to refer to a wide range of actors, from academics, think tank members, and local elites to presidents and prime ministers. There is considerable variation in the term's use 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 the 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 testing theories 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. 2022),

comparisons between different elite samples (Sheffer et al. 2018), designs that use elites for conceptual replications (Kertzer et al. 2021b), to test different sets of micro foundations (Tomz et al. 2020), or as part of a larger design that also incorporates observational data (Hertel-Fernandez et al. 2019). Questions up for debate are 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, such advice focuses mostly on access itself rather than experimental design considerations (a category that includes discussion of how these studies fit into broader research agendas), and a focus on sample size to the exclusion of other factors can miss the point. After all, simpler, smaller experiments can be more compelling than larger, more complex ones if they more directly test 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 fails to accurately simulate the decision-making process it is meant to illuminate. Moreover, the how-to guides available are often tailored toward 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.

## WHO COUNTS AS AN ELITE?

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

Definitional disagreements are critical, since one's conception of eliteness has implications for how elite surveys or experiments ought to be designed as well as how (and if) one can 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. 2022), whereas if we define eliteness in terms of domain-specific expertise, IR 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 an IR context.

There are roughly three different ways in which elites have been conceptualized in the literature, which we categorize as occupational, compositional, and cognitive approaches. Occupational models 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 1976, Bussell 2020). Lasswell (1952), for example, defines elites as "the holders of high positions in a given society" (p. 6) 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

---

**Occupational models of elites:** Models in which political elites are actors whose institutional roles give them influence over public policy

---



---

**Compositional models of elites:**

Models in which political elites are a distinctive class characterized by common socioeconomic characteristics

---

**Cognitive models of elites:**

Models in which political elites are characterized by domain-specific expertise and experience

---

that they happen to occupy unique roles. In contrast, compositional models conceptualize elites as a distinctive class (Meisel 1962) that can be 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 likely to be 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 models, which emphasize elites' possession of 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 than 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 studies, it is not just that elites happen to possess traits like domain-specific expertise, but rather that domain-specific expertise is implied to be 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, as “persons who devote themselves full time to some aspect of politics or public affairs,” a group that includes “politicians, higher-level government officials, journalists, some activists, and many kinds of experts and policy specialists” (Zaller 1992, p. 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 categories. This is particularly true for long-running elite panels—such as the Chicago Council on Global Affairs Leader Survey, which is composed of respondents from the executive and legislative branches, as well as respondents from 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, we note that Renshon (2015) uses a sample drawn from an executive education program at the Harvard Kennedy School along with a matched sample of controls and shows 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 suggest 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 nonelites might sometimes be just as appropriate.

Finally, these divergent conceptions of eliteness probably reflect two factors: variation in the reasons why scholars study elites in the first place as well as practical constraints. In some cases, we survey political elites or conduct experiments on them in order to test theories in which 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. Although all three conceptions of elites are valuable, occupational and compositional conceptions 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, a broader conception of eliteness permits larger samples, more statistical power, and more elaborate designs. 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 heterogeneous samples can sometimes prompt unnecessary confusion. A first step is thus providing a clear conception of what eliteness is for researchers. 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 not 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.
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, transparently describe the composition of the sample.

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 that one has access to affects everything from the recruitment process (best practices for recruiting national security professionals might differ tremendously from recruitment practices targeted toward academics or wealthy citizens) to the design of the study (who 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 matches the theory has commensurate benefits for researchers, as it encourages clearer specification of overall research design considerations, particularly the degree to which the empirical test fits the theory.

For example, experimenters in American politics testing a theory of congressional responsiveness to public opinion by surveying a heterogeneous sample of political elites, which 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. Experimenters in IR testing a theory of presidential decision making by surveying a sample of former national security bureaucrats might note that the sampled elites have domain-relevant expertise, were identified via their former occupation, and are a relatively close match to the theory. These authors should also add 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 inevitably simplifies reality (for example, the absence of time pressure and the nature of the stakes in the scenario).

## 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 narrower conception focused only on when elite studies are particularly valuable, 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.

### When Are Elite Studies Valuable?

Two common rationales offered for elite experiments are substantive justifications and methodological justifications. Substantive justifications center on whether a given study provides further evidence for or against a theory; methodological justifications invoke concepts of external validity, the degree to which a finding extends to other settings or populations (McDermott 2011, p. 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 the *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 (outcomes, treatments, units, and settings) framework (Cronbach & Shapiro 1982, see also Findley et al. 2021). Our discussion below focuses on two of those dimensions: treatments and units.

A first dimension of studies that can help clarify the utility of elites is the experimental treatment itself. We argue that elite studies are particularly informative to the extent that they test theories that directly relate to elites' domain-specific expertise and experience. If, for example, we expect that one of the cognitive advantages of elites is that they make better use of heuristics (Hafner-Burton et al. 2013), it follows that experiments are 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. Nothing in theories of heuristic decision making suggests that elites and nonelites would be differentially likely to use heuristics, so studies extending this research to elites (e.g., Stolwijk & Vis 2020) are motivated primarily by external validity considerations (in a storied tradition of extending behavioral research to unusual samples, e.g., Gigerenzer & Kurzenhaeuser 2005).

Second, elite studies are most useful where the elites being studied (the units) most closely resemble the target population implied by the theory. The use of elite subjects in itself rarely resolves the issue of external validity—which, in any case, is better thought of as a property of research programs and not of 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 it 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 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 only identifying the elites in the theory but also 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, p. 1284)—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. 2022)—it is relatively straightforward to assess how representative the respondents are, due to the presence of accessible benchmarks. For example, biographical and demographic data on legislators are often publicly available, allowing researchers to compare the observable characteristics of legislators who respond to the survey (the sample) with those of the sampling frame as well as the overall population of legislators. In contrast, biographical data are 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 and thereby increasing the likelihood of participation by busy elites. Of course, it also leads to ethical challenges; researchers conducting surveys or experiments on high-ranking elites from well-defined populations often cannot include individual-level demographic covariates in the replication data they publish, lest they inadvertently reidentify respondents.

---

**Complementary experiments:**

Nonoverlapping sets of treatments and outcome measures are administered on both an elite and mass sample

---

## **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 nonexperimental components. Research designs that incorporate elites can be categorized along several dimensions. For projects with multiple experiments, one key feature is whether a common set of treatments is administered across both elite and nonelite samples—allowing for direct comparisons between the two. Other ways of thinking about the aggregation of different methods include whether observational data yield a finding that then motivates an elite experiment, or if observational data are collected after an experiment has been fielded to show similar patterns in different (nonexperimental) 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 is to combine them with other experimental studies conducted either on elites or on mass samples. Some of these studies include paired experiments, where overlapping sets of treatments and outcome measures are administered on both an elite and a mass sample simultaneously, to test whether the patterns observed in a mass sample also hold in an elite one (Kertzer 2022). 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. Other examples are conceptual replications, like that of 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 affect female candidates' electoral success. Another way to incorporate elite studies into larger research designs is complementary experiments, where nonoverlapping sets 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.

While targeted comparisons are useful—both for prosaic purposes of replication and for 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, p. 598), for example, suggest that “scholars should consider investing in elite experiments precisely because research suggests that elites

behave fundamentally differently from nonelites.” Indeed, it is not hard to find evidence suggestive of differences between elite and nonelite samples in paired setups, such as that provided by Mintz et al. (2006), who compared military elites to an undergraduate population in a decision-making experiment. And yet, as Kertzer (2022) points out, eliteness itself is never causally identified in paired studies. This means that scholars sometimes overinterpret differences between elite and mass samples, attributing effects to domain-specific expertise that may instead be a function of compositional differences (e.g., elite samples in many countries typically consist of older, wealthy men). Researchers also sometimes fail to distinguish between differences in intercepts and differences in slopes; they focus 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 combines elite experiments with nonexperimental 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 are used to motivate a problem by drawing attention to variation in political outcomes. For example, Flavin & Hartney’s (2017) analysis of California school board elections finds correlational evidence that school board members are electorally punished only for low achievement levels of white (rather than minority) students in their districts. The authors 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 in German welfare offices, rates of response to German versus 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 researchers combine elite experiments with nonexperimental 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 are measured using publicly available spending and parliamentary attendance data (Ofosu 2019).

## ACCESS AND DESIGN IN ELITE EXPERIMENTS

One of the central challenges of elite experiments relates to design and recruitment. This is particularly true since, as the quantitative literature review shows, scholars fielding elite experiments are more likely to recruit participants themselves than are scholars fielding studies on the mass public. However, the rise of firms like CivicPulse, which 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.

### 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 four 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).
3. Researchers should address the issue of anonymity and duration of the study directly and from the beginning (Goldstein 2002, p. 670; see also Dietrich et al. 2021b) in order to increase 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.

There is less agreement in other areas, such as whether and what type of incentives are appropriate. Dietrich et al. (2021b) suggest monetary incentives, while Godwin (1979) reports that monetary incentives did not increase the sample size of elites responding to a mail survey. Besides, some elites have rules against accepting compensation (Dietrich et al. 2021a) or find monetary incentives insulting (Renshon 2015, p. 674). There is also mixed advice on questions like whether forced-choice designs annoy elites—Maestas et al. (2003) suggest they do, despite Godwin’s (1979) reassurance—and whether to seek approval from leaders of the elite group (e.g., a local chairperson or party leader) prior to fielding the study. Dietrich et al. (2021b) argue that approaching leaders generates “buy-in,” while Vis & Stolwijk (2020, p. 1291) highlight the risk that if the leaders decline to cooperate, the researchers potentially lose a lot of respondents.

A related issue is whether the dangers of “poisoning the well” might be more pronounced with regard to elite subjects than other groups. Political scientists sit uneasily in the middle of a debate on deception, in which psychologists and economists represent opposite poles. To paraphrase McDermott (2013, pp. 605–6), psychologists tend to believe that a little deception isn’t so bad and find that subjects may even prefer participating in research that uses deception (Sharpe et al. 1992), while economists generally prohibit deception, arguing 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 is 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. There are also additional concerns for elites, who, as prominent figures, may worry about being embarrassed in a manner that does not apply to undergraduates. 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 at a sign in the lab that proclaimed it 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 nonmonetary compensation as secondary objects of study in their research. Until such assessments emerge, one is left with a range of helpful anecdotes and experiences worth considering for application to one’s own research. Given the 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 approached 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 recruitment efforts.

---

#### Poisoning the well:

When respondents’ negative experiences from one researcher’s study decrease the likelihood of respondents participating in future studies by other researchers

---

## 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 IR, 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 find little evidence of heterogeneous treatment effects across elite types (for an exception, see Gift & Montan 2021), although many of these subsets are relatively small.

As always, trade-offs 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 samples of elites—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, 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). Like all other methods, then, scholars using elite experiments and surveys confront difficult choices and must balance multiple considerations simultaneously. 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 observational/behavioral data on them is less likely to be available. Heterogeneous samples of elites allow a researcher to bolster a study’s statistical power by obtaining larger sample sizes, but they 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—implies some best practices for our designs. For example, smaller samples suggest the wisdom of (a) more extensive piloting to hone the efficiency and precision of treatments in advance of fielding and (b) the use of pilot data to generate realistic estimates of statistical power. The latter advice is often offered yet rarely followed (Sedlmeier & Gigerenzer 1989). Tools such as DeclareDesign (Blair et al. 2019) are useful in this vein, but their output is only as good as the preparation that goes into them; 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 are 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 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 4–8 times as many subjects as within-subjects designs 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 elites may not be willing to allocate the amount of time necessary for the researcher 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 suggests that elite experimenters need not be overly cautious. Mummolo & Peterson (2019) find little evidence for demand effects in survey experimental designs, freeing researchers to take the advice of Grose (2021, p. 157) to use “meaningful, fairly infrequent and bold interventions” without fear in their elite studies. Another typical concern is that certain subjects might be particularly sensitive for elite respondents, even with assurances of anonymity and nonidentifiable data. In those cases, researchers’ instinct is often to design their treatments in a manner that will not provoke unease or increase attrition; for example, to make vignettes hypothetical or replace the names of countries or actors with fictional names representing types. While some scholars worry that hypothetical questions will generate hypothetical answers, Brutger et al. (2022) show that there are fewer trade-offs associated with abstraction in experimental design than is commonly believed.

## 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 costly, is infeasible, or raises ethical concerns (Nathan & White 2021). Here, two approaches suggest themselves 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 heterogeneous 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 Reserve Officers Training Corps programs, or officers in professional military education programs (e.g., Friedman et al. 2017, Jost et al. 2022); comparativists might expand their aperture to look at former, not just current, legislators—although care should be taken if there have been substantial partisan shifts between legislative terms. If one’s model of eliteness is closer to compositional models that focus on socioeconomic characteristics of elites, then 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 nonelites 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. For example, Tingley (2011) compares the behavior of students who received a lecture on repeated games to that of naive subjects. Other models of eliteness suggest that aspects of elite psychology differ systematically from nonelite psychology, whether those qualities are traits like patience (Hafner-Burton et al. 2014), a 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 trade-offs 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 itself and is appropriate when one's research question implicates features of the setting or context in which actors make decisions. Scholars might ask themselves: does the most important feature to be tested experimentally involve the actors themselves, or rather something about the setting or context in which they make decisions? If the units (elites) are critical, then the options above are available to researchers. If the latter is important, one might build features from the context implicated by the theory into the experimental design. For example, attributes of 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 group dynamics is particularly valuable, given that much elite decision making often takes place in small group rather than individual contexts (Saunders 2017, Kertzer et al. 2022).

## CONCLUSION

In a foundational text in experimental economics, Roth (1986, p. 246) 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 to study them directly, fielding experiments in which political elites themselves are the respondents. An acceleration in the use of elite studies, however, has outpaced our collective understanding of best practices, not only in the logistics of administering elite experiments, but also in the optimal use of elite studies to advance knowledge in the discipline.

In this article, we offer 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 most useful where the elites being studied most closely resemble the target population implied by our theoretical frameworks. However, there are invariably disjunctures between the kinds of elites that researchers are most likely able to access and the ultimate high-level decision makers studied by many of our theories. Being clear about the gaps that arise between our theories and our subjects 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 presented with are as important as the respondents themselves. Recruiting a sample of high-ranking political elites and presenting them with treatments 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 micro foundations of a broader theoretical framework by fielding different experiments on each sample. Researchers can similarly combine elite experiments with nonexperimental 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 nonmonetary forms of compensation actually affect the quantity and quality of elite participation.

Fifth, as with all experiments, researchers fielding elite experiments face trade-offs between sample size, eliteness, and representativeness: e.g., heterogeneous elite samples are typically larger in size but are also more likely to have ill-defined population frames; the more proximate respondents are to positions of political power, the less time they 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, trade-offs remain inevitable, such that researchers should be explicit about how they navigated these trade-offs when designing their study.

Finally, given the challenges of accessing elite samples, scholars without access to elite samples of sufficient size may wish to consider alternative research designs, ranging 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.

## 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:e0193781
- Auerbach AM, Thachil T. 2020. Cultivating clients: reputation, responsiveness, and ethnic indifference in India's slums. *Am. J. Political Sci.* 64:471–87
- Axelrod R. 1976. *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, NJ: Princeton Univ. Press
- Baekgaard M, Christensen J, Dahlmann CM, Mathiasen A, Petersen NBG. 2019. The role of evidence in politics: motivated reasoning and persuasion among politicians. *Br. J. Political Sci.* 49:1117–40
- Bahador B, Entman R, Knüpfer C. 2019. Who's elite and how the answer matters to politics. *Political Commun.* 36:195–202
- Bansak K, Hainmueller J, Hopkins DJ, Yamamoto T. 2018. The number of choice tasks and survey satisficing in conjoint experiments. *Political Anal.* 26:112–19
- Banuri S, Dercon S, Gauri V. 2019. Biased policy professionals. *World Bank Econ. Rev.* 33:310–27
- Bayram AB. 2017. Due deference: cosmopolitan social identity and the psychology of legal obligation in international politics. *Int. Organ.* 71:S137–63

- Bellemare C, Bissonnette L, Kröger S. 2014. *Statistical power of within and between-subjects designs in economic experiments*. Disc. Pap. 8583, Inst. Stud. Labor, Bonn, Ger.
- Blair G, Cooper J, Coppock A, Humphreys M. 2019. Declaring and diagnosing research designs. *Am. Political Sci. Rev.* 113:838–59
- Broockman DE, Ferenstein G, Malhotra N. 2019. Predispositions and the political behavior of American economic elites: evidence from technology entrepreneurs. *Am. J. Political Sci.* 63:212–33
- Broockman DE, Skovron C. 2018. Bias in perceptions of public opinion among political elites. *Am. Political Sci. Rev.* 112:542–63
- Brutger R, Kertzer JD, Renshon J, Tingley D, Weiss CM. 2022. Abstraction and detail in experimental design. *Am. J. Political Sci.* In press
- Busby J, Kafura C, Montén J, Tama J. 2020a. Multilateralism and the use of force: experimental evidence on the views of foreign policy elites. *Foreign Policy Anal.* 16:118–29
- Busby J, Kafura C, Smeltz D, Tama J, Montén J, et al. 2020b. *Coming together or coming apart? Attitudes of foreign policy opinion leaders and the public in the Trump era*. Policy Brief, Chicago Coun. Glob. Aff., Chicago, IL
- Bussell J. 2020. Shadowing as a tool for studying political elites. *Political Anal.* 28:469–86
- Butler DM, Crabtree C. 2021. Audit studies in political science. In *Advances in Experimental Political Science*, ed. JN Druckman, DP Green, pp. 42–55. Cambridge, UK: Cambridge Univ. Press
- Butler DM, Kousser T. 2015. How do public goods providers play public goods games? *Legis. Stud. Q.* 40:211–40
- Butler DM, Nickerson DW. 2011. Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Q. J. Political Sci.* 6:55–83
- Butler DM, Powell EN. 2014. Understanding the party brand: experimental evidence on the role of valence. *J. Politics* 76:492–505
- Carey JM, Helmke G, Nyhan B, Sanders M, Stokes S. 2019. Searching for bright lines in the Trump presidency. *Perspect. Politics* 17:699–718
- Carnes N, Lupu N. 2015. Rethinking the comparative perspective on class and representation: evidence from Latin America. *Am. J. Political Sci.* 59:1–18
- Carnevale JJ, Inbar Y, Lerner JS. 2011. Individual differences in need for cognition and decision-making competence among leaders. *Pers. Individ. Diff.* 51:274–78
- Clifford S, Sheagley G, Piston S. 2021. Increasing precision without altering treatment effects: repeated measures designs in survey experiments. *Am. Political Sci. Rev.* 115:1048–65
- Costa M. 2017. How responsive are political elites? A meta-analysis of experiments on public officials. *J. Exp. Political Sci.* 4:241–54
- Cronbach LJ, Shapiro K. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass
- Dahlberg S. 2007. *Web-based expert surveys*. QoG Work. Pap., Göteborg Univ., Sweden
- Dal Bó E, Finan F, Folke O, Persson T, Rickne J. 2017. Who becomes a politician? *Q. J. Econ.* 132:1877–914
- Dietrich S, Hardt H, Swedlund HJ. 2021a. Elite experiments: strengthening scholarship while bridging the gap. *Duck of Minerva Blog*, June 10. <https://www.duckofminerva.com/2021/06/elite-experiments-strengthening-scholarship-while-bridging-the-gap-2.html>
- Dietrich S, Hardt H, Swedlund HJ. 2021b. How to make elite experiments work in international relations. *Eur. J. Int. Relat.* 27(2):596–621
- Distelhorst G, Hou Y. 2017. Constituency service under nondemocratic rule: evidence from China. *J. Politics* 79:1024–40
- Dynes AM, Hassell HJG, Miles MR, Preese JR. 2021. Personality and gendered selection processes in the political pipeline. *Politics Gender* 17:53–73
- Efrat A. 2015. Cross-national interviewing at international conferences: how to make the most of a unique research opportunity. *Int. Stud. Perspect.* 16:302–11
- Fatas E, Neugebauer T, Tamborero P. 2007. How politicians make decisions: a political choice experiment. *J. Econ.* 92:167–96
- 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. *Int. Organ.* 71:633–63

- Findley MG, Kikuta K, Denly M. 2021. External validity. *Annu. Rev. Political Sci.* 24:365–93
- Findley MG, Laney B, Nielson DL, Sharman JC. 2017b. External validity in parallel global field and survey experiments on anonymous incorporation. *J. Politics* 79:856–72
- Flavin P, Hartney MT. 2017. Racial inequality in democratic accountability: evidence from retrospective voting in local elections. *Am. J. Political Sci.* 61:684–97
- Franco A, Malhotra N, Simonovits G, Zigerell L. 2017. Developing standards for post-stratification weighting in population-based survey experiments. *J. Exp. Political Sci.* 4:161–72
- Friedman JA, Lerner JS, Zeckhauser R. 2017. Behavioral consequences of probabilistic precision: experimental evidence from national security professionals. *Int. Organ.* 71:803–26
- Fuhrmann M, Horowitz MC. 2015. When leaders matter: rebel experience and nuclear proliferation. *J. Politics* 77:72–87
- George AL. 1969. The “operational code”: a neglected approach to the study of political leaders and decision-making. *Int. Stud. Q.* 13:190–222
- George AL, George JL. 1964. *Woodrow Wilson and Colonel House: A Personality Study*. New York: Dover
- Gerring J, Onel E, Morrison K, Pemstein D. 2019. Who rules the world? A portrait of the global leadership class. *Perspect. Politics* 17:1079–97
- Gift T, Monten J. 2021. Who’s out of touch? Media misperception of public opinion on US foreign policy. *Foreign Policy Anal.* 17:128–39
- Gigerenzer G, Kurzenhaeuser S. 2005. Fast and frugal heuristics in medical decision making. In *Science and Medicine in Dialogue: Thinking Through Particulars and Universals*, ed. R Bibace, JD Laird, KL Noller, J Valsiner, pp. 3–15. Westport, CT/Boston: Praeger/Greenwood
- Godwin RK. 1979. The consequences of large monetary incentives in mail surveys of elites. *Public Opin. Q.* 43:378–87
- Goemans HE, Gleditsch KS, Chiozza G. 2009. Introducing Archigos: a dataset of political leaders. *J. Peace Res.* 46:269–83
- Goldstein K. 2002. Getting in the door: sampling and completing elite interviews. *PS Political Sci. Politics* 35:669–72
- Grose CR. 2014. Field experimental work on political institutions. *Annu. Rev. Political Sci.* 17:355–370
- Grose CR. 2021. Experiments, political elites and political institutions. In *Advances in Experimental Political Science*, ed. J Druckman, DP Green, pp. 149–64. New York: Cambridge Univ. Press
- Grossman G, Michelitch K. 2018. Information dissemination, competitive pressure, and politician performance between elections: a field experiment in Uganda. *Am. Political Sci. Rev.* 112:280–301
- Gulzar S. 2021. Who enters politics and why? *Annu. Rev. Political Sci.* 24:253–75
- Hafner-Burton EM, Hughes DA, Victor DG. 2013. The cognitive revolution and the political psychology of elite decision making. *Perspect. Politics* 11:368–86
- Hafner-Burton EM, LeVeck BL, Victor DG, Fowler JH. 2014. Decision maker preferences for international legal cooperation. *Int. Organ.* 68:845–76
- Hafner-Burton EM, Victor DG, LeVeck BL. 2016. How activists perceive the utility of international law. *J. Politics* 78:167–80
- Hemker J, Rink A. 2017. Multiple dimensions of bureaucratic discrimination: evidence from German welfare offices. *Am. J. Political Sci.* 61:786–803
- Hertel-Fernandez A, Mildemberger M, Stokes LC. 2019. Legislative staff and representation in Congress. *Am. Political Sci. Rev.* 113:1–18
- 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 Commun.* 4:1–7
- Hoffmann-Lange U. 2007. Methods of elite research. In *The Oxford Handbook of Political Behavior*, ed. RL Dalton, HD Klingemann, pp. 910–27. Oxford, UK: Oxford Univ. Press
- Hoffmann-Lange U. 2008. Studying elite versus mass opinion. In *The Sage Handbook of Public Opinion Research*, ed. W Donsbach, MW Traugott, pp. 53–63. Los Angeles: Sage
- Humayun F. 2021. *The punisher’s dilemma: domestic opposition and foreign policy losses*. Work. Pap., Yale Univ., New Haven, CT
- Jamison J, Karlan D, Schechter L. 2008. To deceive or not to deceive: the effect of deception on behavior in future laboratory experiments. *J. Econ. Behav. Organ.* 68:477–88

- Jost T, Meshkin K, Schub R. 2022. The character and origins of military attitudes on the use of force. *Int. Stud. Q.* In press
- Karpowitz CF, Monson JQ, Preese JR. 2017. How to elect more women: gender and candidate success in a field experiment. *Am. J. Political Sci.* 61:927–43
- Keller FB. 2016. Moving beyond factions: using social network analysis to uncover patronage networks among Chinese elites. *J. East Asian Stud.* 16:17–41
- Kennedy C, Hartig H. 2019. *Response rates in telephone surveys have resumed their decline*. Pew Res. Cent. <https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/>
- Kertzer JD. 2022. Reassessing elite-public gaps in political behavior. *Am. J. Political Sci.* In press
- Kertzer JD, Busby J, Monten J, Tama J, Kafura C. 2021a. *Elite misperceptions and the domestic politics of conflict*. Work. Pap., Harvard Univ., Cambridge, MA
- Kertzer JD, Holmes M, LeVeck BL, Wayne CN. 2022. Hawkish biases and group decision-making. *Int. Organ.* In press
- Kertzer JD, Renshon J, Yarhi-Milo K. 2021b. How do observers assess resolve? *Br. J. Political Sci.* 51:308–30
- Kertzer JD, Tingley D. 2018. Political psychology in international relations: beyond the paradigms. *Annu. Rev. Political Sci.* 21:319–39
- Krcmaric D, Nelson SC, Roberts A. 2020. Studying leaders and elites: the personal biography approach. *Annu. Rev. Political Sci.* 23:133–51
- Lasswell HD. 1952. The elite concept. In *The Comparative Study of Elites: An Introduction and Bibliography*, ed. HD Lasswell, D Lerner, CE Rothwell, pp. 6–21. Stanford, CA: Stanford Univ. Press
- Leeper TJ, Hobolt SB, Tilley J. 2020. Measuring subgroup preferences in conjoint experiments. *Political Anal.* 28:207–21
- Lin-Greenberg E. 2021. Soldiers, pollsters, and international crises: public opinion and the military's advice on the use of force. *Foreign Policy Anal.* 17:orab009
- Lin-Greenberg E, Pauly R, Schneider J. 2022. Wargaming for political science research. *Eur. J. Int. Relat.* 28:83–109
- Linde J, Vis B. 2017. Do politicians take risks like the rest of us? An experimental test of prospect theory under MPs. *Political Psychol.* 38:101–17
- List JA, Mason CF. 2011. Are CEOs expected utility maximizers? *J. Econom.* 162:114–23
- Loewen PJ, Rubenson D, Wantchekon L. 2010. Help me help you: conducting field experiments with political elites. *Ann. Am. Acad. Political Soc. Sci.* 628:165–75
- Lupia A, McCubbins MD. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York: Cambridge Univ. Press
- Maestas C, Neeley GW, Richardson LE Jr. 2003. The state of surveying legislators: dilemmas and suggestions. *State Politics Policy Q.* 3:90–108
- Mahdavi P. 2017. Scraping public co-occurrences for statistical network analysis of political elites. *Political Sci. Res. Methods* 7:385–92
- Malesky E, Schuler P, Tran A. 2012. The adverse effects of sunshine: a field experiment on legislative transparency in an authoritarian assembly. *Am. Political Sci. Rev.* 106:762–86
- Malhotra N, Monin B, Tomz M. 2019. Does private regulation preempt public regulation? *Am. Political Sci. Rev.* 113:19–37
- McDermott R. 2011. Internal and external validity. The *Cambridge Handbook of Experimental Political Science*, ed. JN Druckman, DP Green, JH Kuklinski, A Lupia, pp. 27–40. New York: Cambridge Univ. Press
- McDermott R. 2013. The ten commandments of experiments. *PS Political Sci. Politics* 46:605–10
- Meisel JH. 1962. *The Myth of the Ruling Class: Gaetano Mosca and the "Elite."* Ann Arbor: Univ. Mich.
- Miller WE, Stokes DE. 1963. Constituency influence in congress. *Am. Political Sci. Rev.* 57:45–56
- 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? *J. Confl. Resolut.* 50:757–76
- Miratrix LW, Sekhon JS, Theodoridis AG, Campos LF. 2018. Worth weighting? How to think about and use weights in survey experiments. *Political Anal.* 26:275–91
- Moyser G, Wagstaffe M. 1987. Studying elites: theoretical and methodological issues. In *Research Methods for Elite Studies*, ed. G Moyser, M Wagstaffe, pp. 1–24. Boston: Allen & Unwin

- Mummolo J, Peterson E. 2019. Demand effects in survey experiments: an empirical assessment. *Am. Political Sci. Rev.* 113:517–29
- Nair D. 2021. “Hanging out” while studying “up”: doing ethnographic fieldwork in international relations. *Int. Stud. Rev.* 23:1399–27
- Nathan NL, White A. 2021. Experiments on and with street-level bureaucrats. In *Handbook of Advances in Experimental Political Science*, ed. J Druckman, DP Green, pp. 509–25. Cambridge, UK: Cambridge Univ. Press
- Neumann IB. 2012. *At Home with the Diplomats: Inside a European Foreign Ministry*. Ithaca, NY: Cornell Univ. Press
- Nyhan B, Reifler J. 2015. The effect of fact-checking on elites: a field experiment on US state legislators. *Am. J. Political Sci.* 59:628–40
- Ofori G. 2019. Do fairer elections increase the responsiveness of politicians? *Am. Political Sci. Rev.* 113:963–79
- Oldendick RW, Bardes BA. 1982. Mass and elite foreign policy opinions. *Public Opin. Q.* 46:368–82
- Ortmann A, Hertwig R. 2002. The costs of deception: evidence from psychology. *Exp. Econ.* 5:111–31
- Page BI, Bartels LM, Seawright J. 2013. Democracy and the policy preferences of wealthy Americans. *Perspect. Politics* 11:51–73
- Palacios-Huerta I, Volij O. 2009. Field centipedes. *Am. Econ. Rev.* 99:1619–35
- Pereira MM. 2021. Understanding and reducing biases in elite beliefs about the electorate. *Am. Political Sci. Rev.* 115:1308–24
- Poole KT, Rosenthal H. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford Univ. Press
- Potters J, van Winden F. 2000. Professionals and students in a lobbying experiment: professional rules of conduct and subject surrogacy. *J. Econ. Behav. Organ.* 43:499–522
- Putnam R. 1976. *The Comparative Study of Political Elites*. Englewood Cliffs, NJ: Prentice-Hall
- Ramey AJ, Klingler JD, Hollibaugh GE. 2016. Measuring elite personality using speech. *Political Sci. Res. Methods* 7:1–22
- Renshon J. 2015. Losing face and sinking costs: experimental evidence on the judgment of political and military leaders. *Int. Organ.* 69:659–95
- Renshon J, Yarhi-Milo K, Kertzer JD. 2022. Democratic reputations in crises and war. *J. Politics*. In press
- Ricart-Huguet J. 2019. The historical turn in the comparative study of political elites. *Newsl. Organ. Sect. Comp. Politics Am. Political Sci. Assoc.* 29:86–93
- Rivera SW, Kozyreva PM, Sarovskii EG. 2002. Interviewing political elites: lessons from Russia. *PS Political Sci. Politics* 35:683–88
- Rodríguez-Teruel J, Daloz JP. 2018. Surveying and observing political elites. In *The Palgrave Handbook of Political Elites*, ed. H Best, J Higley, pp. 93–113. London: Palgrave Macmillan
- Roth AE. 1986. Laboratory experimentation in economics. *Econ. Philos.* 2:245–73
- Saunders EN. 2011. *Leaders at War: How Presidents Shape Military Interventions*. Ithaca, NY: Cornell Univ. Press
- Saunders EN. 2017. No substitute for experience: presidents, advisers, and information in group decision making. *Int. Organ.* 71:S219–47
- Saunders EN. 2022. Elites in the making and breaking of foreign policy. *Annu. Rev. Political Sci.* 25:219–40
- Sedlmeier P, Gigerenzer G. 1989. Do studies of statistical power have an effect on the power of studies? *Psychol. Bull.* 105:309–16
- Shaffer R, Pinson LE, Chu JA, Simmons BA. 2020. Local elected officials’ receptivity to refugee resettlement in the United States. *PNAS* 117:31722–28
- Sharpe D, Adair JG, Roesse NJ. 1992. Twenty years of deception research: a decline in subjects’ trust? *Pers. Soc. Psychol. Bull.* 18:585–90
- Sheffer L, Loewen P. 2019. Electoral confidence, overconfidence, and risky behavior: evidence from a study with elected politicians. *Political Behav.* 41:31–51
- Sheffer L, Loewen PJ, Soroka S, Walgrave S, Sheaffer T. 2018. Nonrepresentative representatives: an experimental study of the decision making of elected politicians. *Am. Political Sci. Rev.* 112:302–21
- 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. *J. Pers. Soc. Psychol.* 110:921–29

- Stolwijk S, Vis B. 2020. Politicians, the representativeness heuristic and decision-making biases. *Political Behav.* 43:1411–32
- Teele DL, Kalla J, Rosenbluth F. 2018. The ties that double bind: social roles and women's underrepresentation in politics. *Am. Political Sci. Rev.* 112:525–41
- 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* 16:792–815
- Tetlock PE, Lebow RN. 2001. Poking counterfactual holes in covering laws: cognitive styles and historical reasoning. *Am. Political Sci. Rev.* 95:829–43
- Tingley DH. 2011. The dark side of the future: an experimental test of commitment problems in bargaining. *Int. Stud. Q.* 55:521–44
- Tomz M, Weeks J, Yarhi-Milo K. 2020. Public opinion and decisions about military force in democracies. *Int. Organ.* 74:119–43
- 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. *Qual. Quantity* 55:1281–317
- Walgrave S, Sevenans J, Van Camp K, Loewen PJ. 2018. What draws politicians' attention? An experimental study of issue framing and its effect on individual political elites. *Political Behav.* 40:547–69
- Yarhi-Milo K. 2018. *Who Fights for Reputation in International Politics? Leaders, Resolve, and the Use of Force*. Princeton, NJ: Princeton Univ. Press
- Yarhi-Milo K, Kertzer JD, Renshon J. 2018. Tying hands, sinking costs, and leader attributes. *J. Confl. Resolut.* 62:2150–79
- Zaller JR. 1992. *The Nature and Origins of Mass Public Opinion*. Cambridge, UK: Cambridge Univ. Press
- Zuckerman H. 1972. Interviewing an ultra-elite. *Public Opin. Q.* 36:159–75