

Re-Assessing Elite-Public Gaps in Political Behavior

Joshua D. Kertzer

Harvard University

Abstract: *Political scientists often criticize psychological approaches to the study of politics on the grounds that many psychological theories were developed on convenience samples of college students or members of the mass public, whereas many of the most important decisions in politics are made by elites, who are presumed to differ systematically from ordinary citizens. This paper proposes an overarching framework for thinking about differences between elites and masses, presenting the results of a meta-analysis of 162 paired treatments from paired experiments on political elites and mass publics, as well as an analysis of 12 waves of historical elite and mass public opinion data on foreign policy issues over a 43 year period. It finds political scientists both overstate the magnitude of elite-public gaps in decision-making, and misunderstand the determinants of elite-public gaps in political attitudes, many of which are due to basic compositional differences rather than to elites' domain-specific expertise.*

Replication Materials: The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/LHOTOK>.

A prominent tradition in political science assumes that systematic differences exist between the behavior of political elites and the mass public. Like the rich in *The Great Gatsby*, elites are not like the rest of us. They have different preferences than the public (Page and Barabas 2000), different value commitments (McClosky 1964), different psychological traits (Hafner-Burton, Hughes, and Victor 2013), and ostensibly even “fundamentally different cognitive architectures” (Kertzer, Renshon and Yarhi-Milo 2019, 18). These elite-public gaps have important implications for democratic theory — what does it mean for political representation when elites neither look nor think like those they govern? — but also for the applicability of political psychology in political science more generally, since many of our theories presume elite cognition not only *differs from* but is also *superior to* that of the public as a whole. The study of leaders in International Relations (IR), for example, has conventionally been about the study of “great

men” (Byman and Pollack 2001), who, blessed with the fruits of expertise, are presumed to be immune from the psychological biases repeatedly appearing in experimental studies of ordinary citizens.

A recent wave of scholarship has sought to subject this assumption about elite exceptionalism to direct empirical testing, fielding paired experiments or surveys on both populations simultaneously. Yet these studies have produced a range of findings that appear difficult to reconcile with one another: some studies find striking similarities between elites and masses (Sheffer et al. 2018), while some point to marked differences (Mintz, Redd, and Vedlitz 2006), and others obtain findings somewhere in between (Renshon 2015). In this article, I seek to make three contributions. First, I offer an overarching conceptual framework to better understand differences between elite and mass cognition. One reason why this recent scholarship has reached such different conclusions is because it has sought to answer a variety of subtly different

Joshua D. Kertzer is Paul Sack Associate Professor of Political Economy, Department of Government, Harvard University. 1737 Cambridge St, Cambridge MA 02138 Web: <http://people.fas.harvard.edu/~jkertzer/> (jkertzer@gov.harvard.edu).

I'm grateful to the authors who generously agreed to share their data, and to Katharina Berninger, Bear Braumoeller, Dan Butler, Devin Caughey, Mike Colaresi, Alex Coppock, Alex Debs, Anita Gohdes, Tyler Jost, Daegyeong Kim, Gary King, Brad LeVeck, Eleonora Mattiacci, Christoph Mikulaschek, Elizabeth Saunders, Lior Sheffer, Nick Smith, Dustin Tingley, Thomas Zeitzoff, the AJPS editors and anonymous reviewers, and audiences at APSA, the Hertie School, Ohio State, the University of Pittsburgh, Yale, and Gareth Nellis' replication seminar at UCSD for helpful feedback and discussions. Special thanks to Max Kuhelj Bugaric for exceptional research assistance. I acknowledge the support of the Weatherhead Center for International Affairs, and the Institute for Quantitative Social Science.

American Journal of Political Science, Vol. 66, No. 3, July 2022, Pp. 539–553

©2020, Midwest Political Science Association.

DOI: 10.1111/ajps.12583

questions, and many of our discussions about elite-mass differences conflate a variety of distinctions (between attitudes, traits, and decision-making) that, I suggest, are better off separated. Elites may have systematically different traits than the public at large, but this doesn't necessarily mean they systematically differ from masses when it comes to decision-making.

Second, I conduct a meta-analysis of paired experiments on political elites and mass samples, comparing 162 treatment effects from 48 paired experiments across 26 studies comparing elites and masses, enabling us to more systematically assess the conditions in which elite and mass decision-making may differ. The meta-analysis shows that even if elites and masses differ in their traits and preferences, they generally respond to treatments in strikingly similar ways. Of the 162 treatment effects I compare between the two sets of samples, only 19 (11.7%) significantly differ in magnitude between elites and masses, and only 3 (1.9%) in sign. At the same time, there are some contexts where elite-public gaps in decision-making are larger than others: a meta-regression finds that the size of these gaps has less to do with the type of political elite being sampled, and more to do with the type of questions being studied, with representation experiments (e.g. Slough 2020; Rosenzweig 2019) displaying the most pronounced gaps.

Third, even in paired experimental studies, it is important to remember that the effect of "eliteness" is not causally identified. Although social scientists often attribute elite-public gaps to elites' domain-specific expertise and experience, there are also often basic compositional differences between the two types of samples, the neglect of which risks mistaking gender or age gaps for expertise effects. Re-analyzing 12 waves of historical public and elite opinion data on foreign policy issues from the Chicago Council on Global Affairs from 1975-2018 (encompassing 1504 polling questions in total, from 5741 elite and 20479 mass public respondents), I show that between a quarter and half of the elite-public gap in foreign policy attitudes in the United States goes away once we account for basic compositional effects. I therefore suggest that political scientists have been both overstating the magnitude and misinterpreting the determinants of elite-public gaps in political behavior.

Elites and Publics

How do political elites differ from the public at large? When is elite political cognition similar to mass political cognition, and when does it differ? These questions

matter for both normative, theoretical, and methodological reasons.

Normatively, these questions matter because of questions of political representation. If politics is governed by a group of elites that looks very different from the constituents they serve (Carnes 2013), or who consistently want different policies than what ruling elites prefer (Page and Bouton 2007), this has important implications for democratic theory, which assumes both that citizens are sufficiently competent to be able to offer guidance to those that govern them, and that those that govern should respond to citizens' wishes.

Theoretically, these questions matter because many of our theories in politics are, if not elitist (Mills 1956; Lupia 2006), then at least keenly interested in bifurcating the study of elite and mass political behavior. Top-down theories of public opinion assume the mass public is "innocent of ideology" (Converse 1964) and knows too little about politics to form meaningful opinions without taking cues from elites (Kertzer and Zeitzoff 2017). In International Relations (IR), many rationalist theories explicitly argue that selection pressures should prevent individuals prone to psychological biases from coming into positions of power. As Copeland (2001, 217) argues, "most leaders, given the obstacles that they must overcome to rise to the top of their nations, do not resemble the 'norm' of the population; that is, we would expect them to be much closer to the ideal-type Machiavellian rationalist than the average citizen." Studying leaders is about studying "great men." (Byman and Pollack 2001) Even in psychological work in IR, Kertzer and Tingley (2018, 327-328) note that "the study of elite political behavior... remains somewhat disconnected from the study of mass political behavior." Scholars traditionally studied each group not just using separate methodological tools, but also distinct theoretical frameworks: operational codes and leadership styles for elites, partisanship and personal values for masses.

Methodologically, these questions matter because one of the common critiques levied against psychological and behavioral research — especially in IR, but also in political science more generally — concerns issues of external validity arising from differences between masses and elites. Historically, much about what we know about psychology comes from a "narrow database" of college students (Sears 1986; Druckman and Kam 2011). In contrast, many of our theories in political science pertain not just to ordinary citizens, but also to political elites: legislators, bureaucrats, generals, diplomats, and so on. For those already unconvinced about psychological work, this disjuncture has long been cause for added skepticism. As Riker (1995, 32) writes, "in decision-making

in political and economic votings, the decision-makers are usually experienced and often are professional elites trained to think about the subject matter of the decision. Experimenters ought to, but usually do not, use trained subjects if they want to convince social scientists of the usefulness of the results.” Even supporters of psychological and experimental work reach similar conclusions. Hafner-Burton, Hughes, and Victor (2013, 368) argue that because elites differ systematically from masses along a host of characteristics (most importantly, domain-specific expertise), “when studying elite decision making it is important to use elite subjects in experimental studies, where possible”, a sentiment echoed by Hardt (2018, 466). Hyde (2015, 406) concurs, characterizing the assumption that “subjects in a lab are like relevant IR populations” as one of the major fault lines in experimental IR. Oberholtzer et al. (2019) go further, arguing that “although observations of players who do not resemble actual decision-makers can produce statistically significant results, they are likely to be irrelevant to real-world policy decisions.”¹

Ultimately, the question of how political elites differ from masses is an empirical one, which as both Sheffer et al. (2018, 304) and Kertzer and Tingley (2018, 328) note, requires using similar methods to study both types of actors; it is difficult to gauge how distinct elite political cognition is from mass political cognition if we only use case studies to study the former and lab experiments to study the latter, for example. Fortunately, the past decade has seen a flurry of studies in which the same methods are applied to both, seeking to subject our assumptions about “elite exceptionalism” (Kertzer 2016, 160) to direct empirical testing. Yet many of these pieces have reported strikingly different findings. Mintz, Redd, and Vedlitz (2006, 770), for example, note that their “results reveal *significant* statistical differences between students and actual decision makers”, while Findley, Milner, and Nielson (2017, 307) note that they find “few substantive differences in citizens’ and elites’ preferences and behavior”. Others offer mixed findings. Hafner-Burton et al. (2014), for example, find that government and business elites are more patient, more strategic reasoners, and more likely to want to join a treaty than college students, but also find that both groups respond similarly to the study’s experimental treatment. Similarly, Linde and Vis (2017, 114)

find that that some of prospect theory’s predictions replicate amongst elites, but not others. How should we reconcile these dramatic differences?

Attitudes, Traits, And Decision-Making

I argue that one of the chief causes of the cacophony of contradictory findings is a lack of conceptual clarity in how we think about elite-public gaps in political behavior. Suppose a standard interactionist framework (e.g. Herrmann, Tetlock, and Visser 1999; Kertzer 2016) of the form:

$$Y = a + B_1X + B_2Z + B_3XZ + \epsilon \quad (1)$$

in which actors’ attitudes, choices, or behavior (Y) are a function of characteristics of the situation the actor faces (B_1X), characteristics of the actor itself (B_2Z), and interactions between the two (B_3XZ). For reasons that will become theoretically important later on, let us decompose Z into two different types of characteristics:

$$Z = B_4\theta + B_5\gamma + \epsilon \quad (2)$$

traits that specifically constitute eliteness (for example, domain-specific expertise and experience – Hafner-Burton, Hughes, and Victor 2013), represented by θ , and all other characteristics that actors might have, represented by γ .

Although deliberately simple, this framework is valuable because it highlights three intuitions. First, there can be a variety of different types of elite-public gaps; one reason why recent scholarship has reached such divergent conclusions is because it has sought to study a range of subtly different phenomena, which political scientists occasionally conflate with one another:

- Elite-public gaps in *attitudes*: differences in what political elites and masses want, or believe to be true. These arise as objects of inquiry both in studies of elite-public gaps in policy preferences (Holsti 2004; Page and Bouton 2007), and in studies of whether elites and publics structure their attitudes in similar ways (Rathbun 2007). These questions are typically answered with observational data. In the econometric framework above, they can be operationalized as differences in Y .
- Elite-public gaps in *traits*: actor-level characteristics that differentiate political elites and masses. For example, recent research has explored whether politicians and voters have

¹This skepticism is likely even more widespread than the above quotations suggest, since it frequently arises in manuscript reviews, which are never published. I am grateful to an anonymous reviewer for this point. For examples of experiments using student samples as proxies for political elites, see Gerber, Morton, and Rietsch (1998); McDermott and Cowden (2001); Tingley and Walter (2011); Kanthak and Woon (2015); Tingley (2017); Lupton (2018).

similar personality traits (Caprara et al. 2003; Dynes et al. 2019), or whether leaders are more self-interested (LeVeck et al. 2014) or more rational (LeVeck 2019). Crucially, these traits are measured both in observational studies — e.g. Dal Bó et al. (2017) study political selection by comparing the cognitive scores of Swedish politicians and general public — and in experimental studies, where they're often used as moderators of treatment effects, as in feelings of power in Renshon (2015), or patience in Hafner-Burton et al. (2014). Whereas individual attitudes might be unstable and highly dependent on context, traits or individual differences are usually understood as capturing general tendencies, which are more stable (Renshon 2008). We can represent these traits with θ .

- Elite-public gaps in *decision-making*: differences in how political elites and masses respond to their environment. These questions are typically answered with paired experiments on elite and mass samples (e.g. Baekgaard et al. 2019; Linde and Vis 2017; Christensen and Moynihan 2020). Respondents are randomly presented with different features or elements in an information environment, and experimenters analyze average differences in responses between treatment conditions. If on average the two groups respond to the treatment differently, it reveals differences in each group's decision-making. In an experimental context, B_1X or $\frac{dY}{dX} = B_1 + B_3Z$ captures decision-making: the effects of features of the environment on choices or behavior, partially as a function of properties of actors themselves.

Thus, many of the mixed findings in the existing literature are due to different pieces of scholarship placing varying degrees of focus on each of these three basic parameters. Sheffer et al. (2018) focus on B_3 and find that elites and masses largely respond to treatments in similar ways; Hafner-Burton et al. (2014) focus both on dispositional traits they associate with eliteness (B_4), and interactions with treatment effects (B_3), finding evidence for elite-public gaps in B_4 , but not in B_3 . The extent to which these gaps are politically consequential depends on which parameter we're interested in, which will be a function of the question we're interested in answering.

Second, because the quantity of interest in most experiments is the average treatment effect (ATE), systematic differences between the traits of elites and masses (θ) only affect decision-making if they interact with the ex-

perimental treatment.² Elites and masses can be characterized by a distinct constellation of dispositional traits, but as long as these traits don't interact with the treatment, these differences will manifest in differing intercepts, rather than the slope of the treatment effect itself.

Third, whether in observational or experimental work, the effect of elite status itself is not causally identified. Because economists' external validity concerns about the first wave of behavioral experiments often focused on questions of expertise and learning (as opposed to "naïve" or "novice" student samples encountering a problem for the first time), much of our extant theoretical apparatus views elites through the prism of domain-specific expertise (e.g. Cooper et al. 1999; Harrison and List 2008; Palacios-Huerta and Volij 2009). This is why Hafner-Burton, Hughes, and Victor (2013) conceptually differentiate political elites and masses by drawing on research in other disciplines discussing, for example, chess grandmasters and musical prodigies. When we see systematic differences between elites and masses in their policy preferences (e.g. Busby et al. 2020), we similarly tend to attribute these differences to elites' relative expertise. Yet it is unclear how analogous political elites really are to Paganini, and in practice, decomposing the effects of Z is difficult; the effect of θ is usually confounded with γ . For example, one frequently encounters media reports gauging President Trump's popularity in the US military, comparing the president's level of popularity amongst troops or veterans versus among the country as a whole.³ One interpretation of this difference is due to θ — individuals who have served in the military have added expertise about the use of force that causes them to support the president at higher rates. Another interpretation is that because military service is nonrandom, there are a variety of basic compositional differences (γ) between samples of military and non-military personnel (Jost, Meshkin, and Schub 2017). The relevant counterfactual when interpreting Trump's popularity among troops therefore involves controlling for the various socioeconomic and ideological variables correlated with both selection into military service and support for Republican presidents.

In the analysis below, I demonstrate each of these points empirically. First, I study elite-public gaps in decision-making using experimental data, conducting a meta-analysis of paired experiments fielded on both

²I am not the first to make a version of this argument; see Druckman and Kam (2011); McDermott (2011).

³See, e.g. <https://www.militarytimes.com/news/pentagon-congress/2018/12/30/poll-shows-high-job-approval-for-trump-from-veterans/>

political elites and mass samples, showing that claims that elites and masses systematically differ in their decision-making are overstated, in that both groups respond to experimental treatments in strikingly similar ways — although there are some issue areas where the gaps are larger than others, most notably political representation experiments requiring one of the samples to guess the beliefs or reactions of the other. Second, I study elite-public gaps in political attitudes using observational data, analyzing historical public and elite opinion data about foreign affairs to show how while it is true that elites and masses differ in their attitudes (as one would expect if they have differing traits), a non-trivial proportion — between a quarter and a half — of these elite-public gaps can be attributed to basic compositional differences between the samples, such that we risk overemphasizing the effect of domain-specific expertise unless we take these more basic differences into account.

Experimental data

Methods

To assess the magnitude of elite-public gaps in decision-making, I turn to experimental data, conducting a meta-analysis of paired experiments on political elites and mass publics. A detailed discussion of the inclusion criteria and data collection procedure is presented in Appendix A (pp. 2-4), but in brief, a study is eligible for the meta-analysis if it contains:

1. an experiment where the treatments are randomly assigned by an experimenter, and
2. the same experiment is fielded both on a sample of political elites (current or former politicians, civil servants, military officers, etc.) and a mass public or convenience sample.⁴

Altogether, the meta-analysis dataset consists of 162 paired treatment effects⁵ that were fielded in 48 paired experiments on samples of political elites and mass publics or convenience samples, from 26 different studies. The first study in the dataset was published in 2006; the most recent were working papers in 2019. Most of these studies were fielded by political scientists, though psychologists and economists are also included here.

⁴See Hafner-Burton, Hughes, and Victor (2013); Bahador, Entman, and Knüpfer (2019) for definitions of political elites.

⁵For purposes of convenience, the paper refers to these throughout as 162 paired treatments, in that the dataset contains 324 treatment effects in total (162 in elite samples, and 162 in mass samples).

The experiments were fielded in 12 different countries (Belgium, Canada, Colombia, Denmark, Israel, Kenya, the Netherlands, Nigeria, Norway, Sweden, Uganda, and the United States). Whereas the mass samples consisted either of student samples, or diverse adult samples (whether non-student convenience samples like participants on Amazon Mechanical Turk, or local or national public samples), the elite samples capture the wide-ranging nature of political elites discussed in this literature: politicians (city councilors, state legislators, current and former national-level politicians, candidates running for elected office), military personnel (either cadets or officers - e.g. Friedman, Lerner, and Zeckhauser 2017; Jost, Meshkin, and Schub 2017), or government bureaucrats and heterogeneous samples of elites. Appendix A (p. 7) shows that the elite samples vary dramatically in size, but are generally smaller than the mass samples, as one might expect; the median sample size for the calculation of each treatment effect is $N=236$ in the elite samples, and $N=976$ in the mass samples.

Substantively, the experiments span American politics, comparative politics, and IR, and involve a wide range of topics and approaches. A quarter of the paired treatment effects come from vignette-based experiments on questions relating to international security or international political economy (e.g. Renshon 2015; Hafner-Burton et al. 2014; Findley et al. 2017), while half of the results come from experiments with questions about domestic politics (e.g. Baekgaard et al. 2019; Malhotra, Monin, and Tomz 2019). Another class of experiments focus specifically on questions of political representation, where the dependent variable involves elites' and masses' beliefs about each others' preferences or behavior (e.g. Rosenzweig 2019). And, a final class of experiments have less explicitly political content, as in stylized bargaining games in the tradition of experimental economics (Butler and Kousser 2015), or studies about research ethics (Naurin and Öhberg 2019). There is thus considerable heterogeneity both in the types of political elites studied in each experiment, as well as the focus of the studies themselves — heterogeneity I model explicitly using meta-regression, below.

Results

For basic notation, let Y_{ijk} denote the response of individual i , in treatment condition $j = \{0, 1\}$, of elite status $k = \{0, 1\}$. Traditionally, the quantities of interest in experiments are average treatment effects ($E[Y_{i10} - Y_{i00}]$, and $E[Y_{i11} - Y_{i01}]$) which leverage the power of random assignment by comparing the difference between

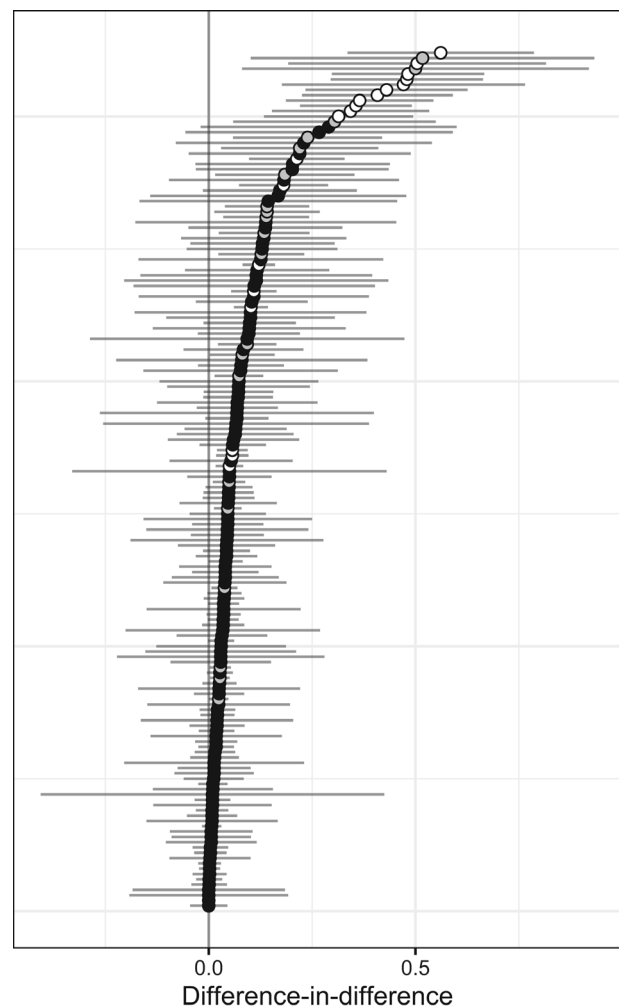
respondents in the treatment condition and respondents in the control condition within each paired comparison. To study elite-public gaps in decision-making, our quantity of interest is the difference-in-difference ($E[Y_{i11} - Y_{i01}] - E[Y_{i10} - Y_{i00}]$): the difference between the treatment and control conditions, for political elites versus masses. Although not formally reported in most of the studies included here, this quantity is useful for our purposes because it tells us whether elites and masses make significantly different decisions as a result of the experimental treatments. Figure 1 presents the difference-in-difference estimates for each paired treatment, with 95% confidence intervals. The plot reveals considerable heterogeneity across studies, confirmed more formally by a χ^2 test from a random-effects meta-analysis with standard errors clustered at the experiment-level ($Q(df = 161) = 379.28, p < 0.001$).⁶ The figure also shows that for the most part, the magnitude of each paired treatment effect doesn't significantly differ between elites and masses; the difference-in-difference estimates are statistically significant in 39 of the 162 cases (24.0%), and if we control the false discovery rate (FDR) using the Benjamini-Hochberg procedure, the estimates are only statistically significant in 19 of the 162 cases (11.7%). These estimates are presented in white in Figure 1.⁷

These difference-in-differences tell us whether the treatment effects significantly differ in *magnitude* between elites and masses, but much of the time we're interested the more basic question of the *direction* of the causal effects. Teele, Kalla, and Rosenbluth (2018), for example, find that local, state, and national legislators in the United States prefer female candidates over male candidates by 7 percentage points, whereas the American public prefers female candidates over male candidates by 4 percentage points; because their elite sample is unusually large, this 3 percentage point difference-in-difference between the two samples is statistically significant. Yet the substantive conclusions the authors draw (that female candidates do not face "outright hostility"), are similar in each case. If, however, one found that female candidates were 1.5 percentage points more popular than male candidates among the public, but 1.5 percentage points *less* popular than male candidates among elites, the difference-in-difference would still be 3 percentage points, but the substantive conclusions we would draw would be rather different. I therefore investigate this

⁶The random effects setup is preferable to a fixed-effects model because it allows the effects to differ across studies, rather than assuming the existence of one true effect. See Riley, Higgins, and Deeks (2011)

⁷See Appendix A (pp. 4-15) for a range of additional tests.

FIGURE 1 Difference-in-Differences between Elites and Masses across Paired Treatments



Note: Figure 1 shows for the most part, the magnitude of each paired treatment effect (with 95% confidence intervals) doesn't significantly differ between elites and masses; the difference-in-difference estimates are not statistically significant at the $p < 0.05$ level in 123 of the 162 cases (75.9%), or in 143 of the 162 cases (88.3%) if we control the false discovery rate (FDR) using the Benjamini-Hochberg procedure. Estimates statistically significant after controlling for multiple comparisons are shown in white; estimates statistically significant only as long as multiple comparisons are not controlled for are shown in grey

additional question directly, estimating the average treatment effects within each sample for each paired treatment effect, identifying those ATEs that (i) significantly differ from zero in both samples, and (ii) differ from one another in sign. Of the 162 paired treatment effects in the data, only 3 (1.9%) significantly differ in sign between elites and masses.

That is to say, although experimentalists and non-experimentalists alike frequently express skepticism

about experiments testing theories of elite decision-making on non-elite samples, the treatment effects recovered in the elite samples included in this analysis do not significantly differ in magnitude from those recovered from mass samples 88% of the time, and do not significantly differ in sign 98% of the time. Alternately, it is possible that the relative absence of elite-mass differences in decision-making across these paired experiments speaks to a different type of external validity concern, revealing the limitations of the experiments themselves: if the experiments had more mundane realism, or involved tasks or context more closely relating to what the political elites do as part of their profession, perhaps the effects of expertise would kick in (Cooper et al. 1999, 786; Levitt, List, and Reiley 2010, 1414).

Explaining Heterogeneity Across Experiments. Since there is significant heterogeneity across the results in Figure 1, we can model this heterogeneity formally using a meta-regression model, a form of meta-analysis which utilizes covariates to help explain the variation uncovered between studies.⁸ Meta-regressions are particularly valuable because they let us control for multiple sources of heterogeneity simultaneously, in a way that estimating separate versions of Figure 1 for different types of elite samples, for example, cannot.⁹ In Table 1 I estimate a mixed-effects model with standard errors clustered at the experiment level, studying the effect of four types of experiment-level characteristics in particular.

To capture the heterogeneous contexts in which the experiments were fielded — one might imagine elite-public gaps would be starker in developing countries than industrialized ones, for example — I use GDP per capita data from the International Monetary Fund, although supplementary models in Appendix A (pp. 12–15) include seven other country-level covariates, from country size to individualism-collectivism. In all cases, I fail to find evidence that these country-level characteristics significantly explain variation across studies.

To model the diversity of elites included across experiments, I include a set of dichotomous variables for politicians (either current or former, at either the local, national, or subnational level), military personnel (either cadets, or officers), and bureaucrats/non-elected policymakers. Table 1 shows there's some evidence that

TABLE 1 Metaregression Analysis of the Magnitude of Elite-Public Gaps

	β	SE	<i>p</i>
Intercept	−0.088	0.101	0.400
Contextual factors			
GDP per capita	0.001	0.001	0.686
Elite sample characteristics			
Military	−0.024	0.031	0.470
Politician	0.079	0.034	0.049
Mass sample characteristics			
Student	0.125	0.038	0.011
Experimental domain			
Representation	0.217	0.094	0.042
Domestic politics	0.050	0.024	0.118
IPE	0.026	0.075	0.738
International security	0.106	0.037	0.030

Note: Random-effects meta-regression with clustered SEs at the experiment-level shows elite-public gaps are significantly larger in representation experiments. Reference categories for elite sample characteristics: bureaucrats or heterogeneous elite samples; for experimental domain: apolitical tasks (e.g. stylized experimental economic games).

compared to non-elected bureaucrats, politicians tend to display slightly larger elite-public gaps — contrary to the presumption that elected officials have more of an incentive to think like their constituents — but the substantive effect is small.

Relatedly, since the magnitude of the elite-public gap may be a function not just of the type of elite being studied, but also the public sample to which the elites are being compared, I include a dichotomous variable indicating whether the mass sample is a convenience sample of students (either at the undergraduate level, or higher), rather than a diverse adult sample (either at the community- or national-level). Table 1 shows that elite-mass studies that use student samples obtain significantly larger elite-public gaps than those that use more diverse adult samples. However, supplementary analysis in Appendix A (p. 15) suggests this result is being driven by a significant interaction effect between the type of elite and public sample: elite-public gaps are particularly pronounced when the elite sample consists of military officials, and the public sample consists of undergraduates.

Finally, since the experiments vary notably in their topic or focus, I include a set of dichotomous variables indicating whether the studies focus on questions in international security, international political economy, domestic politics, representation, or an apolitical reference category. The presence of this last category is particularly important, because differences between

⁸For other recent meta-analyses in political science, see Costa (2017); Kalla and Broockman (2018).

⁹Certain types of elites are more likely to be studied in certain types of countries, in regards to certain types of questions, such that simply comparing differences between elites and masses among elected politicians, versus non-elected elites, for example, would lead to biased inferences.

political elites and masses should be larger in political domains than nonpolitical ones (McDermott 2011, 27), where elites' domain-specific expertise is less relevant. Table 1 shows that compared to this apolitical reference group, our elite-mass gaps are larger in experiments about international security issues, but are especially large in representation experiments. Indeed, many of the paired treatments for which we see the biggest elite-public gaps in Figure 1 share something in common: seven of the thirteen largest difference-in-differences come from studies of political representation in developing countries, where the dependent variables involve elites' and masses' beliefs about the behavior of masses, or elites' and masses' beliefs about the behavior of elites. Rosenzweig (2019) finds Kenyan voters dislike politicians who use violent electoral strategies more than Kenyan politicians assume, while Slough (2020) finds that Colombian and Nigerian citizens expect disadvantaged citizens will receive relatively less favorable service from bureaucrats than Colombian and Nigerian bureaucrats expect. In these studies, elite and mass samples differ from one another less because of systematic differences in decision-making, but rather, because of potential misperceptions by elites (see also Broockman and Skovron 2018; Hertel-Fernandez, Mildenerger, and Stokes 2019).

Observational data

Method

The experimental results above focus on elite-public gaps in *decision-making*: to what extent, and in which circumstances, do elites and publics respond differently to information presented in experimental treatments? To assess the magnitude of elite-public gaps in *political attitudes* — and the proportion of these gaps attributable to basic compositional effects, rather than domain-specific expertise — I turn to observational data, compiling a set of a historical public opinion polls on American public and elite opinion in foreign policy issues originally fielded by the Chicago Council on Foreign Relations (later the Chicago Council on Global Affairs), every few years beginning in 1975. The elite sample consists of “foreign policy opinion leaders”: a heterogeneous elite sample drawn from Congress, the executive branch, foreign policy think tanks, interest groups, and academics.¹⁰ The mass pub-

lic sample consists of nationally representative samples of the American public, originally fielded by telephone by Gallup, and eventually online by YouGov.

This data is ideal for our purposes for four reasons. First, the low levels of knowledge most Americans have about international affairs makes foreign policy an “easy case” for the existence of elite-public gaps, and one where domain-specific expertise should be all the more consequential. Second, the data have unusual breadth. Rather than just studying elite-public gaps in a single case or with a small handful of questions — as is the case in most experimental studies — the data here cover a diverse array of topics across a wide span of time: the dataset I build below consists of 1504 individually-matched foreign policy questions, from 26220 respondents (5741 foreign policy elites, and 20479 members of the public) across twelve waves from 1975 to 2018.¹¹ The questions in the data thus range widely across both topic and time, from whether the military should play a more important role in American foreign policy, to support for foreign aid; from whether Henry Kissinger places too much of an emphasis on secret diplomacy, to whether the US should have a long-term military base in Guantanamo Bay.¹²

Third, samples of political elites, especially at high levels of government in a country like the United States, are relatively difficult to obtain today, which is one reason why many elite experiments on elected officials draw heavily on local or provincial politicians, who are easier to access than national-level ones.¹³ Because the Chicago Council studies began in an era when political elites weren't inundated with requests to participate in academic studies, the data here include access to unusually high-level respondents: as I show in greater detail in Appendix B (pp. 27–30), the respondents from Congress in 1986, for example, were sampled from members (“beginning with chairmen and ranking members”) from House and Senate committees and subcommittees specially focused on foreign affairs, such as the House

¹¹Similar to the paired treatment effects in the experimental section, each of these 1504 questions appears both in a public survey and an elite survey fielded that same year.

¹²Importantly, these studies, and the analysis below, compare elites' preferences on a given policy issue with the public's preferences on that issue. The outcome variable and quantity of interest here thus differs from studies like Broockman and Skovron (2018), which are interested instead in comparing the public's preferences with elites' *perceptions* of the public's preferences on a given issue.

¹³In the meta-analysis in the previous section, 89 of the 162 paired-treatment effects come from studies of politicians; 55 of the 89 (62%) include local or provincial politicians. None of the studies where the politicians sampled are exclusively at the national-level were fielded in the United States.

¹⁰For more details about the sample composition, and robustness checks showing the results hold regardless of whether the analysis below is conducted on the full sample versus the subsample of elites from the legislative and executive branches, see Appendix B (pp. 27–30).

Committee on Foreign Affairs, the Senate Committee on Foreign Relations, the Defense and Foreign Operations Subcommittees of the House and Senate Committees on Appropriations, and so on — in an era when these committees exerted more oversight, and their members had far greater foreign policy expertise, than today (Goldgeier and Saunders 2018). The domain-specific expertise reflected in the elite sample — directly connected to the policy questions on which the elite-public gaps are being measured — makes it an especially fruitful test of the role of domain-specific expertise in political cognition.

Finally, this data is also valuable because classic studies by Page and Barabas (2000) and Page and Bouton (2007) used Chicago Council polls to document a “foreign policy disconnect” between leaders and masses. Since this previous work was interested in the contours of political representation, it focused much of its analysis on the specific policy issues where the gaps between elites and masses are the largest: for example, elites’ lower levels of isolationism and protectionism than the public at large. Since my interest is in assessing the overall magnitude of elite-public gaps, however, I look instead at the full set of questions, analyzing them in the aggregate.¹⁴

Results

I begin by rescaling each question to range from 0-1 to facilitate direct comparability, and present the raw distribution of elite and mass foreign policy preferences by year in Figure 2. Figure 2 shows that while there is a moderate level of agreement between mass preferences (on the x axis) and elite preferences (on the y axis), there are also marked elite-public gaps much as Page and Bouton (2007) reported, as illustrated by the loess smoother, and the dispersion of the points in each panel. It also shows that the magnitude of these gaps varies by year. In 2016, the correlation between elite and mass preferences was only $r = 0.38$; in 1986, the correlation between elite and mass preferences was $r = 0.80$.

Yet these two groups differ from one another in a variety of ways beyond elites having domain-specific foreign policy experience. While demographic data are missing from some of the early elite surveys, the demographic information available (detailed in Appendix B (pp. 18-21)) shows stark compositional differences be-

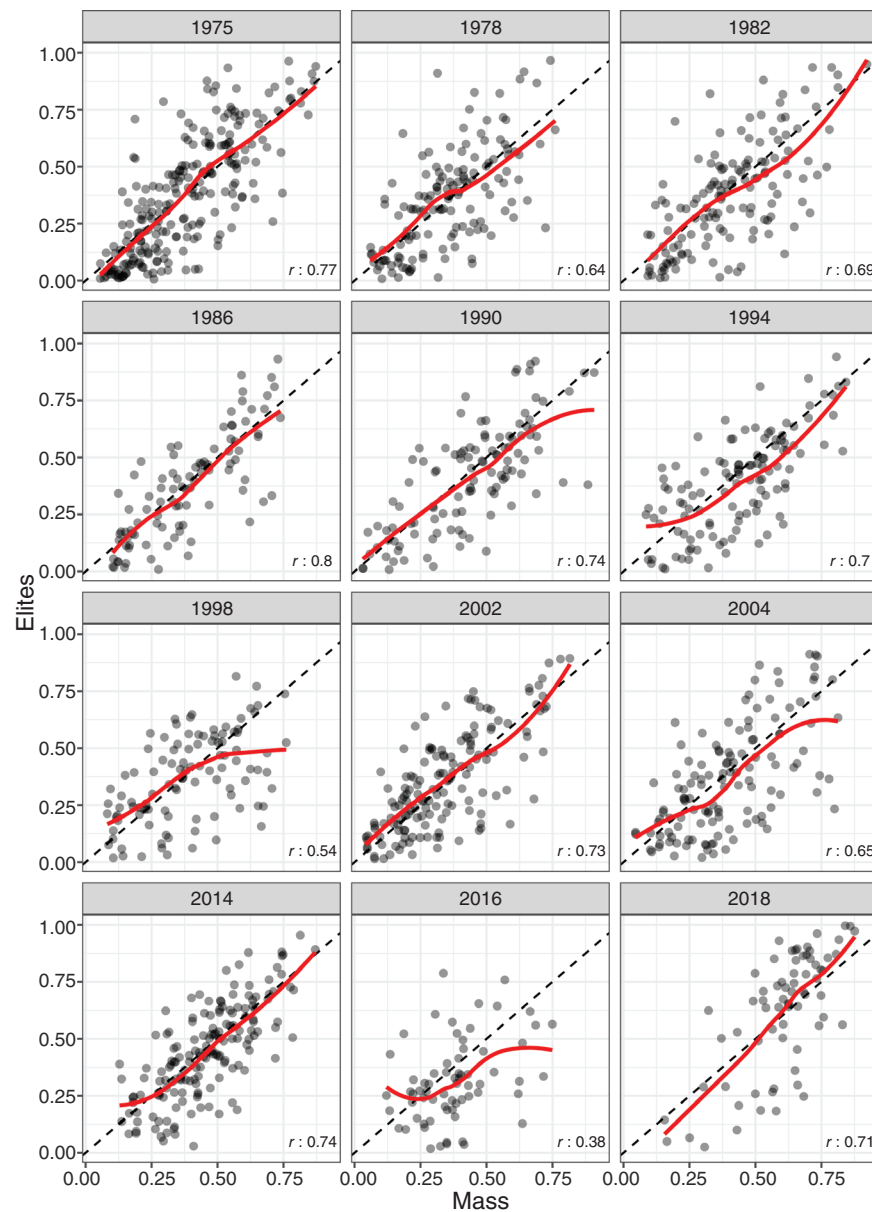
tween the two samples. For example, whereas men made up roughly 50% of the mass public sample in each year, they made up 97% of the elite sample in 1975, 94% of sampled elites in 1986, and 81-83% of sampled elites from 1998-2014, before reaching an all-time low of 74% of the elite sample in 2018. Education levels in the elite sample are only available in 1998 and 2014-2018, but in these four waves, 24-35% of the mass sample had a four-year college degree, compared to 93-99% of the elite sample. Income data is not available for the elite sample, but given what we know elsewhere about the socioeconomic background of political elites (Carnes 2013), one can presume our elite sample is probably wealthier on average than the public sample as a whole. The age data available suggests the elite sample tends to be older than the public, especially in the early waves of the study. Finally, Appendix B (pp. 18-21) presents mosaic plots for age distributions across the two samples for those years in which data is available. The plots show that in 1975 and 1998, the elite sample mostly consists of 40-64 year olds, thereby skewing older than the public as a whole. By the most recent surveys, the age composition of the elite sample has shifted, perhaps reflecting survey mode differences as the elite surveys went online.

On the one hand, these compositional differences presumably accurately characterize demographic differences between political elites and masses; it should not be surprising to most observers of contemporary politics or scholars of descriptive representation that American political elites are highly-educated and historically have been more likely to be male, for example (Barnes and Holman 2019; Gerring et al. 2019). On the other hand, however, these gaps between elites’ and masses’ policy preferences are often used to point to the centrality of domain-specific expertise in elite psychology; to borrow from the theoretical framework presented earlier, these gaps are attributed to differences in θ , rather than γ . Yet given the amount of research documenting marked gender gaps in public opinion about foreign policy (Mansfield, Mutz, and Silver 2015; Eichenberg 2016), or generational differences in foreign policy preferences (Holsti 2004), simply comparing elite and public preferences in foreign affairs without accounting for these demographic differences conflates an elite-public gap with a gender gap, or age gap. These gender and age gaps may be important for questions of political representation, but are usually not what political scientists have in mind when they talk about elite-public gaps in foreign policy preferences.

I adopt a variety of empirical strategies (including Blinder-Oaxaca decomposition analysis) in Appendix B (pp. 22-27), but for the analysis in the main text

¹⁴The analysis includes all foreign policy attitude question from these surveys with ordered response options; I exclude questions with unordered responses because of the difficulty in directly comparing elite and mass responses to them in an aggregate way without relying on subjective modeling decisions.

FIGURE 2 Elite and Mass Foreign Policy Preferences, by Year

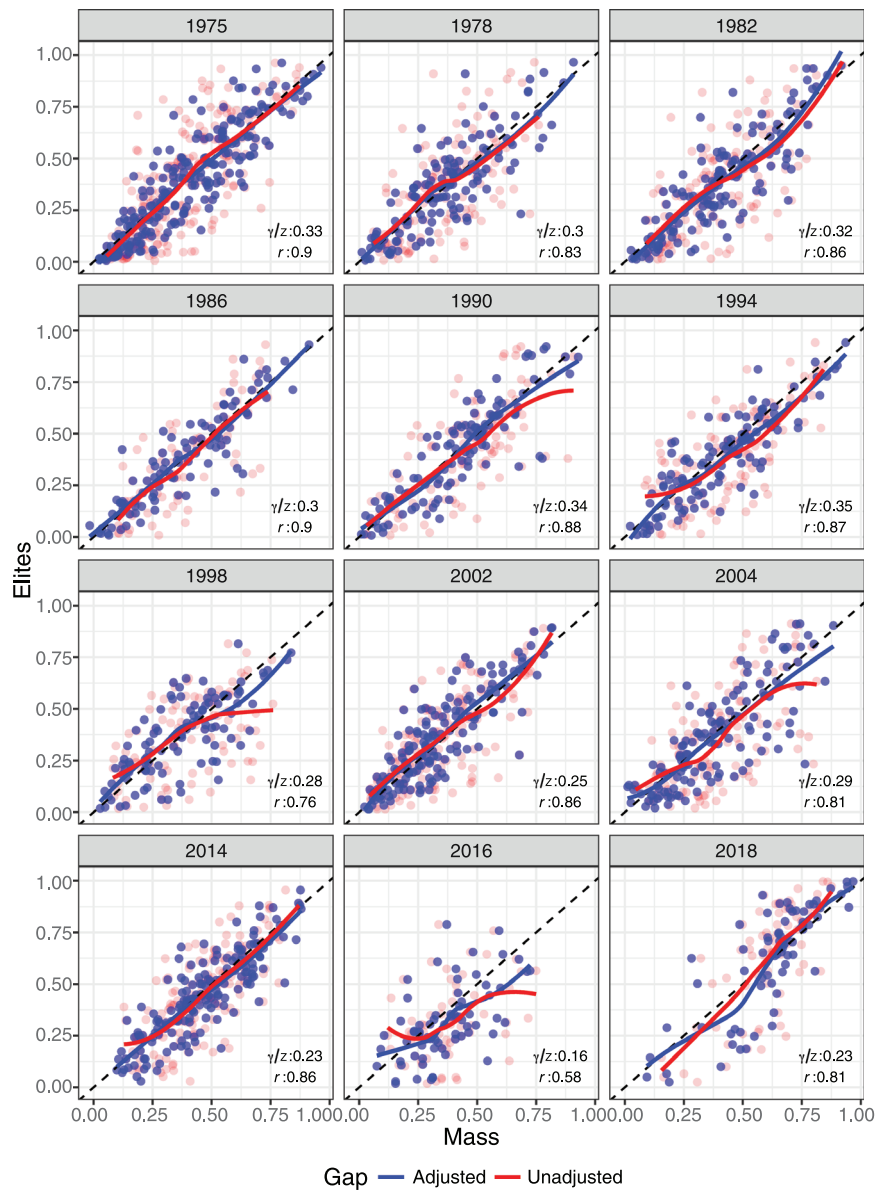


Note: Each point represents elite and mass responses to a different polling question in a given survey wave. The closer each point is to the 45° diagonal line, the more agreement there is between elites and masses on the issue. The correlation coefficients (r) show that for most years, elite and mass foreign policy preferences are well correlated, but there are some issues that feature marked disagreements

I adopt a simpler approach, which requires making fewer assumptions about the data. For each wave of the survey, I resample the public data with replacement $B = 5000$ times; in each bootstrap, I estimate a series of linear regression models of the form $y_{ijk}^* = \hat{\beta}_0 + \hat{\beta}_1 \text{College} + \hat{\beta}_2 \text{Male} + \hat{\beta}_3 \text{Income}_{50} + \hat{\beta}_4 \text{Age}_{40} + \epsilon$, in which each of the foreign policy questions

$j = \{1 \dots n\}$ asked in year $k = \{1975, \dots, 2018\}$ is regressed on a series of dummy variables denoting whether the respondent has a college degree, is male, has an above-median income for the public sample in that year, and is aged 40 or above. I then save the predicted values ($\bar{Y}_{jk}^* = \sum_{b=1}^B \frac{y_{ijk}^*}{B}$), thereby obtaining an estimate of what a counterfactual public sample composed of

FIGURE 3 Elite and Mass Foreign Policy Preferences, after Adjusting for Compositional Differences



Note: The light dots replicate Figure 2, whereas the dark dots depict elite and mass foreign policy preferences after adjusting the demographic composition of the mass sample. The dark dots cluster closer to the 45° diagonal, showing that the adjusted public attitudes generally more closely resembles the elite attitudes, confirmed by the higher correlation coefficients (r). The γ/Z estimates in each panel indicate the proportion of the elite-public gap in a given year attributable to demographic differences, suggesting that apart from 2016, between 23 and 35% of the elite-public gap in foreign policy attitudes can be attributed to mundane compositional differences between the elite and mass samples

40+ year-old males with college degrees and above-median incomes thinks about each foreign policy issue.¹⁵

¹⁵See Appendix B pp. 22-26 for supplementary analysis using a subsetting method instead to avoid linearity assumptions, which obtains similar results.

Figure 3 depicts the results of these simulations, presenting adjusted public preferences alongside the unadjusted preferences from Figure 2. Despite the bluntness of the empirical strategy, the adjusted public attitudes generally more closely resemble the elite attitudes, clustering closer to the 45° diagonal. The correlation coefficients in

each panel show that apart from 2016, the correlation between elite and public attitudes markedly improves after adjustment, ranging from $r = 0.76$ in 1998 to $r = .90$ in 1986. $\frac{\gamma}{\Sigma}$ estimates the proportion of the elite-public gap in political attitudes in a given year that can be attributed to basic demographic differences — in other words, to γ rather than θ . The plot suggests that apart from 2016, between 23 and 35% of the elite-public gap can be attributed to mundane compositional differences between the elite and mass samples.

Four points here are worth noting. First, the clear exception to the trends noted above is 2016, where the correlation after adjustment remains relatively low ($r = 0.58$), and only 16% of the elite-public gap is attributable to basic demographic differences; 2018 performs slightly better ($r = 0.81$; $\frac{\gamma}{\Sigma} = 0.23$), but is also lower than many of the other years. While it is possible this shift is due to the rise of Donald Trump, supplementary analyses in Appendix B (pp. 25–26) suggest that this pattern is likely due to changes in the partisan composition of the elite sample, which becomes significantly more Democratic-leaning than the public sample does in this same time period. When we replicate the analysis above, but also adjusting for partisanship, 40% of the elite-public gap in 2016 and 49% of the gap in 2018 is attributable to demographic differences, and the correlation between elite-public preferences jumps up to $r = 0.79$ and $r = 0.95$, respectively. Second, supplementary analysis in Appendix B (pp. 25–27) using two-way Blinder-Oaxaca decomposition analysis suggest the simulation approach likely underestimates the proportion of elite-public gaps attributable to compositional differences, such that the results reported above are likely conservative estimates.

Third, as noted above, the analysis here relies on the complete sample of foreign policy leaders, a heterogeneous elite sample that includes elected politicians, but also represents other quadrants of the foreign policy establishment: special interest groups, thinktanks, academics, and so on. Appendix B (pp. 27–30) repeats the analysis in the main text, but restricting the elite sample under consideration to members of the executive and legislative branch (who on the one hand, may be more “elite”, but on the other, may be more incentivized to think like the public due to electoral pressures), finding strikingly similar results as those reported above.¹⁶

Fourth, my claim is not that domain-specific expertise and experience are irrelevant. Although basic

compositional differences make up to half of the elite-public gap in attitudes in this data, there is still another half of the elite-public gap that remains unexplained, some — though not all — of which may be due to expertise.¹⁷ What this analysis suggests, however, is that viewing elite political behavior solely through the prism of expertise causes us to miss out on many more banal reasons why we might see elite-public gaps in political attitudes.

Conclusion

In this article, I argued that political scientists have both overstated the origins and misunderstood the consequences of elite-public gaps in political behavior. I did so while seeking to make three contributions. First, I offered an overarching conceptual framework, enumerating three different features on which elites and publics may differ: their attitudes, their traits, and their decision-making. Many of the mixed findings in the existing literature are due to different scholars focusing on each of these questions. Second, to explore differences in decision-making, I conducted a meta-analysis of 162 paired treatment effects from paired experiments on political elites and masses. The results showed that although elites and masses may differ in their traits, this does not necessarily mean they will significantly differ in their decision-making; in the studies included here, the treatment effects recovered in the elite samples did not significantly differ in magnitude from those recovered from mass samples 88% of the time, and did not significantly differ in sign 98% of the time. The studies where we see the largest differences are studies of political representation, where the two samples differ from one another not because of systematic differences in decision-making, but rather because of misperceptions by elites.

These findings suggest political scientists’ reflexive skepticism about experiments conducted on non-elite samples — and thus, the applicability of psychological theories to the study of political elites — may be unwarranted. It should also invite further study about the theoretically relevant traits that differentiate elites and masses. A growing body of literature on how personality traits shape selection into politics, for example, shows

¹⁶See Appendix B (pp. 30–31) for a discussion of how elite cues might affect the results.

¹⁷For example, elites and publics may express different foreign policy preferences because they differentially bear the impacts of a particular policy: elites may be more interventionist and less protectionist than the public not because of expertise or experience, but because they are less likely to bear the costs of war than the public is, and more likely to be the “winners” from free trade.

that a non-random sample of individuals choose to run for higher office (e.g. Dynes, Hassell, and Miles 2019; Clifford, Kirkland, and Simas 2019; Dynes et al. 2019). These traits may not only help us theoretically better understand the origins of elite-public gaps, but point to important moderators of treatment effects in paired experiments. They also suggest fruitful lines of future research, such as meta-analyses testing contextual variation of trait differences between elites and masses, or longitudinal studies testing the conditions in which expertise contributes to these gaps (Jost, Meshkin, and Schub 2017).

Third, although we tend to view elites through the prism of domain-specific expertise and experience, the effect of eliteness in these studies is confounded with a host of basic demographic characteristics that have little to do with domain-specific experience. Analyzing 1504 polling questions from 12 waves of historic public opinion data on foreign policy over a 43 year period, I showed that elite-public gaps in foreign policy attitudes are real, but misunderstood, since between a quarter and a half of these gaps can be attributed to basic demographic differences. From a normative perspective, these elite-public gaps are important regardless of whether they come from θ or γ . But from a theoretical perspective, this distinction between the two is highly consequential: the findings here suggests as much as half of these elite-mass differences stem not from the superiority of elite cognition, but because political scientists have been mistaking a gender gap, or an age gap, or an income gap, with a gap in expertise.

The findings thus also have implications for the study of public opinion in foreign policy more generally: classical realists like Kennan and Morgenthau argued that foreign policy should be insulated from the whims of the mass public because experts correctly perceive the national interest to an extent that novices do not. These findings suggest many of these persistent elite-public differences in foreign policy are due to factors more mundane than domain-specific expertise. They also emphasize the value of descriptive representation in foreign policy: while some of the IR literature argues that female leaders behave similarly to male ones when in power (Horowitz, Stam, and Ellis 2015), the analysis above suggests that up to half of the “foreign policy disconnect” between Washington and the country as a whole has to do with the former not looking as much like the latter as it might otherwise. Together, these findings raise deeper questions about the nature of political expertise, and suggest caution about claims that the cognition of political elites markedly differ from ordinary citizens.

References

- Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen. 2019. “The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians.” *British Journal of Political Science* 49(3):1117–40.
- Bahador, Babak, Robert Entman, and Curd Knüpfer. 2019. “Who’s Elite and How the Answer Matters to Politics.” *Political Communication* 36(1):195–202.
- Barnes, Tiffany D, and Mirya R Holman. 2019. “Taking diverse backgrounds into account in studies of political ambition and representation.” *Politics, Groups, and Identities* 7(4):829–41.
- Broockman, David E., and Christopher Skovron. 2018. “Bias in Perceptions of Public Opinion Among American Political Elites.” *American Political Science Review* 112(3):542–63.
- Busby, Joshua, Craig Kafura, Jonathan Monten, and Jordan Tama. 2020. “Multilateralism and the Use of Force: Experimental Evidence on the Views of Foreign Policy Elites.” *Foreign Policy Analysis* 16(1):118–29.
- Butler, Daniel M, and Thad Kousser. 2015. “How do public goods providers play public goods games?” *Legislative Studies Quarterly* 40(2):211–40.
- Byman, Daniel L., and Kenneth M. Pollack. 2001. “Let Us Now Praise Great Men: Bringing the Statesman Back In.” *International Security* 25(4):107–46.
- Caprara, Gianvittorio, Claudio Barbaranelli, Chiara Consiglio, Laura Picconi, and Philip G. Zimbardo. 2003. “Personalities of Politicians and Voters: Unique and Synergistic Relationships.” *Journal of Personality and Social Psychology* 84(4):849–56.
- Carnes, Nicholas. 2013. *White-Collar Government: The Hidden Role of Class in Economic Policy Making*. Chicago, IL: University of Chicago Press.
- Christensen, Julian, and Donald P. Moynihan. 2020. “Motivated reasoning and policy information: Politicians are more resistant to debiasing interventions than the general public.” *Behavioural Public Policy* Forthcoming.
- Clifford, Scott, Justin H. Kirkland, and Elizabeth N. Simas. 2019. “How Dispositional Empathy Influences Political Ambition.” *Journal of Politics* 81(3):1043–56.
- Converse, Philip E. 1964. The nature and origin of belief systems in mass publics. In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press, pp. 206–61.
- Cooper, David J., John H. Kagel, Wei Lo, and Qing Liang Gu. 1999. “Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers.” *American Economic Review* 89(4):781–804.
- Copeland, Dale C. 2001. “Theory and history in the study of major war.” *Security Studies* 10(4):212–39.
- Costa, Mia. 2017. “How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials.” *Journal of Experimental Political Science* 4(3):241–54.
- Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. “Who Becomes A Politician?” *Quarterly Journal of Economics* 132(4):1877–1914.

- Druckman, James N., and Cindy D. Kam. 2011. Students as Experimental Participants: A Defense of the 'Narrow Data Base'. In *Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. Cambridge: Cambridge University Press, pp. 41–57.
- Dynes, Adam M., Hans J. G. Hassell, and Matthew R. Miles. 2019. "The Personality of the Politically Ambitious." *Political Behavior* 41(2):309–36.
- Dynes, Adam M., Hans J. G. Hassell, Matthew R. Miles, and Jessica Robinson Preese. 2019. "Personality and Gendered Selection Processes in the Political Pipeline." *Politics & Gender* Forthcoming:1–21.
- Eichenberg, Richard C. 2016. "Gender Difference in American Public Opinion on the Use of Military Force, 1982–2013." *International Studies Quarterly* 60(1):138–48.
- Findley, Michael G., Adam S. Harris, Helen V. Milner, and Daniel L. Nielson. 2017. "Who Controls Foreign Aid? Elite versus Public Perceptions of Donor Influence in Aid-Dependent Uganda." *International Organization* 71(4):633–63.
- Findley, Michael G., Helen V. Milner, and Daniel L. Nielson. 2017. "The Choice among Aid Donors: The Effects of Multilateral vs. Bilateral Aid on Recipient Behavioral Support." *Review of International Organizations* 12(2):307–34.
- Friedman, Jeffrey A., Jennifer S. Lerner, and Richard Zeckhauser. 2017. "Behavioral Consequences of Probabilistic Precision: Experimental Evidence from National Security Professionals." *International Organization* 71(4):803–26.
- Gerber, Elisabeth R., Rebecca B. Morton, and Thomas A. Rietz. 1998. "Minority Representation in Multimember Districts." *American Political Science Review* 92(1):127–44.
- Gerring, John, Erzen Oncel, Kevin Morrison, and Daniel Pemstein. 2019. "Who Rules the World? A Portrait of the Global Leadership Class." *Perspectives on Politics* 17(4):1079–97.
- Goldgeier, James, and Elizabeth N. Saunders. 2018. "The Unconstrained Presidency: Checks and Balances Eroded Long Before Trump." *Foreign Affairs* 97(5):144–56.
- Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor, and James H. Fowler. 2014. "Decision Maker Preferences for International Legal Cooperation." *International Organization* 68(4):845–76.
- Hafner-Burton, Emilie M., D. Alex Hughes, and David G. Victor. 2013. "The Cognitive Revolution and the Political Psychology of Elite Decision Making." *Perspectives on Politics* 11(2):368–86.
- Hardt, Heidi. 2018. "Who matters for memory: Sources of institutional memory in international organization crisis management." *Review of International Organizations* 13(3):457–82.
- Harrison, Glenn W., and John A. List. 2008. "Naturally Occurring Markets and Exogenous Laboratory Experiments: A Case Study of the Winner's Curse." *The Economic Journal* 118(528):822–43.
- Herrmann, Richard K., Philip E. Tetlock, and Penny S. Visser. 1999. "Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93(3):553–73.
- Hertel-Fernandez, Alexander, Matto Mildemberger, and Leah C. Stokes. 2019. "Legislative staff and representation in congress." *American Political Science Review* 113(1):1–18.
- Holsti, Ole R. 2004. *Public Opinion and American Foreign Policy*. Revised edition ed. Ann Arbor, MI: University of Michigan Press.
- Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. 2015. *Why Leaders Fight*. New York, NY: Cambridge University Press.
- Hyde, Susan D. 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18:403–24.
- Jost, Tyler, Kaine Meshkin, and Robert Schub. 2017. "Socialized Hawks? How Selection Explains Military Attitudes on the Use of Force." Working paper: Last accessed August 6, 2020. http://www.tylerjost.com/uploads/1/1/0/4/110425699/socializedhawks_full.pdf
- Kalla, Joshua L., and David E. Broockman. 2018. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 112(1):148–66.
- Kanthak, Kristin, and Jonathan Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." *American Journal of Political Science* 59(3):595–612.
- Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton, NJ: Princeton University Press.
- Kertzer, Joshua D., and Dustin Tingley. 2018. "Political Psychology in International Relations: Beyond the Paradigms." *Annual Review of Political Science* 21:319–39.
- Kertzer, Joshua D., Jonathan Renshon, and Keren Yarhi-Milo. 2019. "How Do Observers Assess Resolve?" *British Journal of Political Science* Forthcoming.
- Kertzer, Joshua D., and Thomas Zeitoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3):543–58.
- LeVeck, Brad L. 2019. "Mutual belief in rationality among international policy elites." <https://convention2.allacademic.com/one/apsa/apsa19/>. APSA 2019 Conference Paper. Last accessed August 6, 2020.
- LeVeck, Brad L., D. Alex Hughes, James H. Fowler, Emilie M. Hafner-Burton, and David G. Victor. 2014. "The Role of Self-Interest in Elite Bargaining." *Proceedings of the National Academy of Sciences* 111(52):18536–41.
- Levitt, Steven D., John A. List, and David H. Reiley. 2010. "What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments." *Econometrica* 78(4):1413–34.
- Linde, Jona, and Barbara Vis. 2017. "Do Politicians Take Risks Like the Rest of Us? An Experimental Test of Prospect Theory Under MPs." *Political Psychology* 38(1):101–17.
- Lupia, Arthur. 2006. "How elitism undermines the study of voter competence." *Critical Review* 18(1-3):217–32.
- Lupton, Danielle L. 2018. "Signaling Resolve: Leaders, Reputations, and the Importance of Early Interactions." *International Interactions* 44(1):59–87.
- Malhotra, Neil, Benoit Monin, and Michael Tomz. 2019. "Does Private Regulation Preempt Public Regulation?" *American Political Science Review* 113(1):19–37.

- Mansfield, Edward D., Diana C. Mutz, and Laura R. Silver. 2015. "Men, Women, Trade, and Free Markets." *International Studies Quarterly* 59(2):303–15.
- McClosky, Herbert. 1964. "Consensus and Ideology in American Politics." *American Political Science Review* 58(2):361–82.
- McDermott, Rose. 2011. Internal and External Validity. In *Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge University Press, pp. 27–40.
- McDermott, Rose, and Jonathan A. Cowden. 2001. "The effects of uncertainty and sex in a crisis simulation game." *International Interactions* 27(4):353–80.
- Mills, C. Wright. 1956. *The Power Elite*. Oxford University Press.
- Mintz, Alex, Steven B. Redd, and Arnold Vedlitz. 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):757–76.
- Naurin, Elin, and Patrik Öhberg. 2019. "Ethics in Elite Experiments: A Perspective of Officials and Voters." *British Journal of Political Science* Forthcoming.
- Oberholtzer, Jenny, Abby Doll, David Frelinger, Karl Mueller, and Stacie Pettyjohn. 2019. "Applying wargames to real-world policies." *Science* 363(6434):1406–06.
- Page, Benjamin I., and Jason Barabas. 2000. "Foreign Policy Gaps between Citizens and Leaders." *International Studies Quarterly* 44(3):339–64.
- Page, Benjamin I., and Marshall M. Bouton. 2007. *The Foreign Policy Disconnect: What Americans want from our leaders but don't get*. Chicago, IL: University of Chicago Press.
- Palacios-Huerta, Ignacio, and Oscar Volij. 2009. "Field Centipedes." *American Economic Review* 99(4):1619–35.
- Rathbun, Brian C. 2007. "Hierarchy and Community at Home and Abroad: Evidence of a Common Structure of Domestic and Foreign Policy Beliefs in American Elites." *Journal of Conflict Resolution* 51(3):379–407.
- Renshon, Jonathan. 2008. "Stability and Change in Belief Systems: The Operational Code of George W. Bush." *Journal of Conflict Resolution* 52(6):820–49.
- Renshon, Jonathan. 2015. "Losing Face and Sinking Costs: Experimental Evidence on the Judgment of Political and Military Leaders." *International Organization* 69(3): 659–95.
- Riker, William H. 1995. "The Political Psychology of Rational Choice Theory." *Political Psychology* 16(1):23–44.
- Riley, Richard D., Julian PT Higgins, and Jonathan J Deeks. 2011. "Interpretation of random effects meta-analyses." *BMJ* 342:d549.
- Rosenzweig, Steven C. 2019. "Dangerous Disconnect: Voter Backlash, Elite Misperception, and the Costs of Violence as an Electoral Tactic." Working paper: Last accessed August 6, 2020. https://www.dropbox.com/s/dyopen95ufeiei/Dangerous_Disconnect_Article__May_16_2019.pdf?dl=0
- Sears, David O. 1986. "College Sophomores in the Laboratory: Influence of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51(3):515–30.
- Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave, and Tamir Sheafer. 2018. "Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians." *American Political Science Review* 112(2):302–21.
- Slough, Tara. 2020. "Bureaucrats Driving Inequality in Access: Experimental Evidence from Colombia." Working paper: Last accessed August 6, 2020. http://taraslough.com/assets/pdf/colombia_audit.pdf
- Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112(3):525–41.
- Tingley, Dustin. 2017. "Rising Power on the Mind." *International Organization* 71(S1):S165–88.
- Tingley, Dustin H., and Barbara F. Walter. 2011. "The Effect of Repeated Play on Reputation Building: An Experimental Approach." *International Organization* 65(2):343–65.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Experimental data

Appendix B: Observational data