Methodologies for “Political Science as Problem Solving”

Cyrus Samii

New York University

January 20, 2023

1 Introduction

I am going to offer a framework, which we can call “political science as problem solving,” that can bring resolution to the following questions:

- How can the empirical research of political scientists be useful?
- How do we know whether there is a strong ethical basis for our research?
- How might descriptive, observational-causal, and interventionist/experimental-causal empirical research relate?
- How can theory motivate empirical research?

High-level methodological questions like these are often swimming around in the minds of researchers and students trying to find their groove. I appreciate that not all political scientists are worried about these questions, but to the extent that one is, a problem-solving approach is clarifying. If you buy into this idea, then it also has implications for how methodological training should be organized.

I will begin by explaining what I mean by a problem-solving approach and arguments that justify it. The key is to motivate research in terms of questions like “what is the problem here?", "why does the problem persist?", and "how can we mitigate the problem?" By "problem" I mean something that is *normatively wrong* in the world. This requires normative commitments.

I will then proceed to discuss the three stages of development in a problem-solving research program: establishing the problem, establishing mechanisms, and testing interventions. Each stage is distinctive methodologically. In stage 1, inference is *descriptive*, in stage 2 it is *observational causal*, and in stage 3 it is *interventionist causal.* Each section is organized by first explaining goals for the given stage of the research program and then skills needed to pursue those goals. When it comes to describing empirical skills, I will focus on quantitative methods, and point to useful texts and examples for political scientists. I do not address qualitative methods here because they our beyond my areas of expertise. However I would note that qualitative evidence most certainly plays an important role in each of the three stages discussed below. The characterization of the three stages and the core skills that I identify could help to organize sequences for quantitative methods training.

2 A Problem-Solving Approach

"Political science as problem solving" conceives of research programs as collective efforts to address normatively-defined, real-world problems. The approach resembles strategies from management, such as the Six Sigma "Define, Measure, Analyze, Improve, Control" (DMAIC) approach (De Mast and Lokkerbol, 2012). But this idea has also been proposed as a basis for social science research programs by others more eloquent than me. Moynihan (2022) draws on Simon (2019) to propose that scholars of political science and public administration can find purpose by appreciating their roles in helping to *design* institutions that better serve normative goals. As Moynihan (2022, p. 1) puts it, social scientists have a special role here:

> Public organizations lack market pressures for improvement. Democratic forces for change may miss the target, focused on political goals, some of which may

> themselves be antidemocratic or prioritize dysfunction over performance. Hierarchically imposed solutions, especially in authoritarian regimes, often lead to disastrous outcomes, as they miss the necessary feedback loops and experimentation needed for government to succeed. The applied study of government, properly understood, is able to not just bear witness to such problems but play a role in resolving them.

Duflo (2017) argues that social scientists have methodological and substantive expertise that is especially useful for designing real-world policy. Deft handling of theory and accumulated evidence contributes to good policy design, deep understanding of human behavior contributes to good implementation, and methodological expertise contributes to learning about policy impact. Along with other social scientists, political scientists have something to contribute to real-world problem solving. If you find value in doing so, then you should.

There is a sequence to the development of a problem-solving research program. First is a persuasive problem definition. This involves the interplay of normative analysis (a particular fact pattern *would be* a problem, in principle) and empirical description (the troubling fact pattern is indeed apparent). The relevant empirical methods here are primarily *descriptive*.

Second is using evidence and theory to establish persuasively *mechanisms* that are important drivers of the problem. By mechanisms I have in mind what Cowen (1998, p. 127) defines as "processes through which initial conditions operate through human behavior to produce a final result." The relevant empirical methods here are primarily *observational-causal*, seeking to explain naturally occurring phenomena. Given that we are ultimately interested in problem *solving*, breakthrough findings are not be defined simply in terms of how much variation is persuasively *explained*, but rather whether the mechanism identified is one that would appear to be *amenable* to intervention.

Third is the design and execution of persuasive tests of compelling interventions to disrupt relevant mechanisms so as to, hopefully, mitigate the problem. The relevant empirical methods here are primarily *interventionist-causal*, seeking to test an intervention that may be wholly novel to the world. Interventionist-causal analyses are often in the form of experiments, although the term "experiment" can refer to approaches to measurement that are more descriptive (e.g., conjoint "experiments" are not interventionist-causal, but rather descriptive). Moreover interventionist-causal analyses may sometimes rely on quasi-experimental methods, in situations where randomization was either not feasible or not the approach used by whomever guided the intervention. In the current discussion, the "observational vs. experimental" distinction is less important than a distinction between observational analyses seeking to *explain naturally occurring* phenomena versus interventionist analyses seeking to *test interventions*.

I use the term "persuasive" in defining the three steps of development of a problem-solving research program. This is meant as a reminder about the social nature of social science research. It also reflects the fact that problem-solving research programs operate as appeals for how attention and resources should be applied to try to improve the world. Typically such decisions require appeal to collective interests and overcoming skeptics.

Often times graduate students are taught to think of research in terms of solving puzzles. A compelling theory may suggest that things should go one way (e.g., Downsian convergence for electoral candidate platforms) but then observation deviates, yielding a puzzle. Unintended consequences can also yield puzzles. Grofman (2001, p. 1), in a volume entitled *Political Science as Puzzle Solving*, proposes that attention to "concrete puzzles deriving from empirical observation" serves as a meaningful basis for theorizing and empirical research. I see the puzzle-solving approach as relating to the problem-solving approach in two ways. First, step 2 of the developmental sequence described above (explaining naturally occurring phenomena) involves puzzle solving, at least some of the times. Second, problem-solving motivation provides a basis for establishing the importance of a puzzle. For example, Golden (2001)'s contribution to the volume edited by Grofman strikes us as meaningful insofar as it contributes to explanations for erosion of labor rights, a normatively-defined problem.

In the sections that follow, my aim is to propose how the problem-solving approach draws from different areas of quantitative methodology.

3 Establishing the Problem

The first stage of a problem-solving research program is to establish, descriptively, that there is "something wrong in the world" that deserves attention and resources to try to address. The goal is to assess the importance of the problem and persuade skeptics (whether those disinclined to believe that the problem *matters*, or those disinclined to believe that it *doesn't*) about your assessment of its importance. Doing so involves skills, including normative argument to motivate the relevant concepts and the problem, the operationalization of measures to characterize the problem, and then compelling empirical description on the basis of this operationalization. Normative arguments are not my emphasis here, and so I will not discuss methods of normative argumentation. I will just offer two points. First, normative political theory has a crucial role to play, and yet it seems all too rare that normative political theorists and empirical researchers come into direct conversation with each other to define real-world problems. At the very least, empirical researchers should take this as a plea to start reflecting on how their work connects with normative concerns (e.g., Kymlicka, 2002). Second, political science is different from classical economics in that the latter focuses on Pareto or Kaldor-Hicks efficiency, which by definition does not require defending the value of interventions that ultimately benefit one group relative to others. By contrast, generic normative criteria that political scientists emphasize, like civic equality, may motivate interventions intended to improve the standing of some individuals relative to others. This difference is often the basis of defining something as "political."

My emphasis here will be on core skills in empirical methodology—from operationalization to measurement to description. Given a problem that is defined persuasively in normative terms, one wants to investigate the importance of the problem in terms of extent or severity. One may want to examine whether things are trending in a positive or negative direction. Problems may be defined in terms of differences or disparities between groups. The nature of the empirical work here is primarily *descriptive*. Quantitative description is a broad and very rich subdomain of political science methodology drawing on statistical theory, causal inference, analysis of strategic interactions, programming, and visualization.

One must begin by defining the population within which the problem resides and the measures that one could use to operationalize the concepts that are implicated in the problem. Good measures can track improvement or exacerbation of the problem. Different measures may track short-term change, but perhaps be more susceptible to error, or longer-term change. Thus, the first core set of skills are those used to assess the validity, reliability, and comparability of measures (Zeller and Carmines, 1980; Adcock and Collier, 2001; Schedler, 2012). This set of skills can be used in methodological work on improving existing measures (e.g., King, Murray, Salomon, and Tandon, 2004) or in finding ways to link measures picking up on short-term variation to variation in slower-to-change outcomes of interest (e.g., Athey, Chetty, Imbens, and Kang, 2019). Variables that make sense, conceptually, for characterizing a problem are sometimes directly observable or accessible through respondents' self-reports. But they may also be latent or indirectly revealed by observed choice behavior. Measurement of latent quantities is thus an important area of descriptive methodology in political science. Political scientist have developed a rich array of tools to measure policy preferences, whether by scaling observed choice behavior, such as roll-call votes, (Laver, 2014), scaling media content (Ho, Quinn, et al., 2008), designing survey-based choice tasks (Sniderman, 2018; Bansak, Hainmueller, Hopkins, and Yamamoto, 2021), or otherwise designing strategies for survey respondents to reveal attitudes or behaviors that may be sensitive (Blair, Imai, and Zhou, 2015). Political scientists are also contributing to methods for measuring latent social structures, such as social networks (Mahdavi, 2019; Bisbee and Larson, 2017).

Once measures are defined, another core skill area is in constructing data. This could come from scraping administrative or digital sources or from surveys or other field-based observation. Political scientists are increasingly contributing to strategies for extracting, linking, and otherwise processing administrative data to create analysis-ready datasets (e.g., Enamorado, Fifield, and Imai, 2019). Design of descriptive field research strategies, including survey sampling, is a longstanding domain of expertise among political scientists (e.g., Weisberg, 2009), although it is too rarely taught within political science graduate programs.

Another core skill set for descriptive inference is the ability to diagnose sources of bias or misrepresentation owing to selection processes, misreporting, or strategic dynamics. Knox, Lowe, and Mummolo (2020) and Westwood, Grimmer, Tyler, and Nall (2022) each demonstrate how a modern causal inference framework can help to diagnose biases in what are intended to be descriptive analyses, respectively, of racial bias in police shootings and survey respondents' support for political violence. Consideration of strategic dynamics can inform how well different fact patterns represent the scale or extent of a problem. Consider the problem of restrictions on individuals' ability to exercise their right to vote. If a particular group is subject to greater restrictions, then in the first instance it might be that the problem should be evident in turnout numbers. But if the intention to vote is sufficiently strong, then the burden imposed by the restrictions may not be revealed strongly by turnout numbers, and rather outcomes such as time and effort put into voting (Pettigrew, 2017). The restrictions could induce countermobilization to sustain turnout levels, although at a cost that could be avoided if the restrictions were removed (Cantoni and Pons, 2021). This example helps to show how strategic analysis can help in interpreting how descriptive patterns relate to the scale or extent of a problem.

Finally, compelling presentation of data so as to reveal important trends or population conditions is another core skill in descriptive inference. Visualization is often a compelling way to characterize the extent or severity of a problem, e.g., by constructing trend lines or plots that show how outcomes vary over a population. Textbooks such as Yau (2011) and Healy (2018) are excellent for providing inspiration and techniques. That said, data visualizations do not have to be fancy to be compelling, they just need to be designed in a way that allows us to view relevant levels, comparisons, or trends. Presentation and critique of visual presentations of data should be a regular part of discussions of empirical research (Gelman, 2010).

Thus, to recap, problem-solving political science draws on the following skill sets in helping to establish problems:

- Normatively defining problems,
- Operationalizing the concepts used in the problem statement and defining measures that are valid, reliable, and comparable,
- Constructing data by scraping, linking, sampling, etc.,
- Diagnosing sources of bias or misrepresentation, and
- Presentation of trends or patterns that reveal the extent and severity of the problem, including visualizations.

The newly established *Journal of Quantitative Information* reflects renewed attention among political scientists to many of the methodological areas described above. That said, contributions to methods for measurement and characterization of societal problems are of quite general interest, as is apparent from the fact that many of the contributions referenced above appear in our discipline's top general interest journals.

4 Establishing Mechanisms

Having established, descriptively, that a problem is substantial, the next stage of a problem-solving research program is to try to establish mechanisms that perpetuate the problem. This involves the theorizing of mechanisms, deriving observable implications and hypotheses, and *observational causal* empirical research to test these hypotheses so as to assess the relevance of potential mechanisms. The goal here is, on the one hand, diagnostic: finding a causal process through which the problem is perpetuated. However, such diagnosis is not, in itself, the end-goal. The end goal is to define potential interventions. Thus, establishing mechanisms that are *amenable to intervention* or have *programmatic implications* is especially important because it allows one to move on to the intervention stage. This goal of intervention gives priority to some types of explanations over others. The mechanisms that are amenable to intervention may not be the things that explain the most variation.

Turning to relevant skills, most of what political science students learn as being "quantitative methodology" corresponds to this activity of testing observable implications of mechanisms, although it is not always (or perhaps even often) the case that this is understood to be with the aim of finding mechanisms that could be intervened-upon to address normatively-defined problems. Whether oriented toward problem-solving or not, the emphasis in this stage is in characterizing causal relations in the naturally-occurring processes that determine variation in the outcomes of interest.

This first key skill that political scientists would have to apply in this stage is in using analytical methodologies (like game theory or behavioral theory) to propose mechanisms. Again, my emphasis here is on empirical methods, and so I will not discuss methods for building positive theories. Like what I proposed above about how normative theorists and empirical researchers could do more to find connections across each other's work, an emphasis on problem solving can be a basis on which positive theorists and empirical researchers find stronger connections across each others' work. Empirical researchers need not be theorists themselves, they rather need to consult theory in motivating mechanisms to test empirically.

Given a theory that yields mechanism-type propositions, a second set of skills is in characterizing observational implications of mechanisms. Mechanisms can be represented with "directed acyclic graphs" (DAGs) that relate the key causal factors that a mechanism highlights to the outcomes of interest (outcomes that define the problem), along with an accompanying narrative account to explain the relations that the DAG represents between the key causal factors and the outcomes (Pearl and Mackenzie, 2018). A good research design at this stage is one that can reliably estimate the strength of relationships implied by the DAG so as to assess whether the DAG is capturing first-order important mechanisms. An important skill, although one that does not appear to be taught formally in existing programs, is translating theories to DAGs (Rodrigues, Kreif, Lawrence-Jones, Barahona, and Mayer, 2022). This is an area of metholodology that deserves much more attention. Political scientists should also be well-versed in deducing the relationships between variables that a DAG implies and in relating empirical strategies to manipulations on a DAG (Greenland and Pearl, 2014). Such relationships include the causal effects that the DAG represents between the key causal factors and the outcomes. Such relationships also include non-causal relationships between variables that might be *induced* by the causal relationships captured by the DAG and that depend on naturally occurring selection or choice processes in the real world. The next few sections explain.

Capturing the causal effects that a DAG represents requires skill in using observational research designs to establish causal identification for the effects of key causal factors and also to establish ways to estimate other relationships implied by the DAG (e.g., moderation or mediation effects). This is the focus of much instruction in contemporary political science graduate programs, drawing on seminal work by Angrist and Pischke (2009), Imbens and Rubin (2015), Pearl and Mackenzie (2018), Morgan and Winship (2015), and Rosenbaum (2010), Ashworth, Berry, and de Mesquita (2021), among others. A fundamental skill area is in mastering identification strategies, such as deriving covariate control strategies from a DAG, recognizing useful instrumental variables or discontinuities, or understanding the identifying power of over-time variation. Statistical inference for causal effects is also a crucial skill area (Abadie, Athey, Imbens, and Wooldridge, 2020).

Causal identification analysis can also involve detecting more subtle problems with research designs. Aronow and Samii (2016) and Samii (2016) discuss situations in which identification strategies relying on covariate control, such as multiple regression, can yield estimates for "effective samples" that may depart substantially from intended target populations. It is also well-understood that instrumental variables that induce exogenous variation in treatment variables (i.e., causal factors of interest) identify "local effects" for compliers—

that is, units whose treatment status is affected by the instrument (Angrist, Imbens, and Rubin, 1996). De Mesquita and Tyson (2020) discuss another way that the sources of exogenous variation can create "commensurability problems", in situations where the causal factors of interest are behaviors that can convey information about those engaging in the behavior. For example, the effect of a protest that emerges as the result of opportunities that a government cannot observe may differ from the effects of a protest in which opportunities are observable, due to differences in the inferences that the government would make about the types of people protesting. Regression discontinuity designs also identify local effects for those close to the relevant cut-off. The literature on school quality notes that regression discontinuity studies of the effect of admission to selective schools may be confounded by "rank effects": those barely admitted on the basis of test scores may be among the lowest ranking students in the selective school, potentially negating the benefits that the selective school offers (Delaney and Devereux, 2022). Marshall (2022) shows how using regression discontinuity designs to estimate the effects of *characteristics* of elected officials can be problematic. Suppose one wanted to estimate the effect of electing women as representatives, and one compared places in which a women barely won to places in which a woman barely lost. The challenge here is that the women and men may differ in their respective qualities that lead them to end up in close races. Then, the analysis is confounded by the existence of such "compensating differentials" across men and women.

Other than estimating causal effects that a DAG represents, one can also assess the relevance of a mechanism by assessing the strength of non-causal, *induced* relationships that the associated DAG implies. Such induced relationships result from situations in which what we observe in the real world is somehow based on conditioning on "collider" variables in a DAG (Pearl and Mackenzie, 2018). Identification analysis usually raises the issue of conditioning on colliders when trying to analyze sources of selection bias (Elwert and Winship, 2014). But conditioning on colliders can occur through natural selection and choice processes, inducing statistical relationships that can be revealing as to the causal process at work (Becker, 1993).

Among those using formal theory, such non-causal induced relationships are sometimes referred to as relationships "at equilibrium," in which case what we see is conditional on the rational best-response choices of agents under study. Ashworth et al. (2021, pp. 19-28) and Cohen (2021) analyze the problem of women's underrepresentation in elected office along these lines. If women tend to underrate their abilities to serve in office, then conditional on running for office, women candidates would tend to possess higher ability than male candidates. This has implications for how men and women candidates are likely to perform electorally and legislatively.

To what extent are experiments useful for evaluating mechanisms in an explanatory sense? Many survey or lab experiments, especially, appear to be motivated by a desire to generate causal identification for explanatory purposes (that is, to test theories), rather than as attempts to test interventions aimed at mitigating real world problems. Such "theory-testing" experiments are different from, e.g., survey experiments like audit, list, endorsement, or conjoint experiments or lab-in-the-field experiments that are more about characterizing, in a descriptive sense, individual preferences, biases, or attitudes. Generally speaking, I am wary about using experiments primarily to test theories. If the experiment is done in a very controlled and fabricated setting (a lab with students as subjects, or as a survey experiment), one must make a leap to establish a relationship with real-world mechanisms. If the experiment is done as a field experiment in the real world, then to the extent that one is really affecting people's lives, I believe that the experimenter has an ethical obligation for the experiment to operate with the primary intention of trying to mitigate real world problems, rather than just using other people's lives for the purpose of academic inquiry. Experiments operating with a problem-solving intention are part of the third "interventionist" stage of the problem-solving research program, and are meant to be informed by research under this second "observational" stage. Of course researchers can draw theoretical insights from tests of interventions, but to set that as the only goal is problematic in the basis of a "do no harm"-type ethics.

An example may help to clarify the logic here. Consider a research program that is motivated by the question of why members of majority groups do not support efforts to protect rights that are defined generally but have intensified implications for minority group members. Suppose a theory hypothesizes that such non-contribution is motivated by the social distance that majority group members feel toward the minority group. A theory-motivated experiment might test this hypothesis by randomly assigning a treatment that *increases* this feeling of social distance, and then assesses whether this depresses support for effort to protect the rights in question. From an academic perspective, this would seem to make sense, and many experimental studies, particularly survey experimental studies, are set up like this. I find such an experiment to be problematic. The treatment pushes individuals into an counterfactual reality that we would not want to create if our interest was in addressing the problem of majority non-support. If the intervention has effects that linger beyond the experiment itself, it might make the problem worse. I would be much more comfortable with a research program that proceeded by *observationally* investigating majority group members' sense of social distance toward minority group members and the effects of such perceived social distance. If the observational evidence is compelling, then this warrants designing an intervention that tries to improve things by *reducing* perceived social distance and evaluating effects on support for efforts to protect rights. We establish prevailing levels of social distance, and all that it implies, as the control condition, and the counterfactual is one that generates a change that we have theoretical reason to believe is beneficial. The intervention is not motivated merely to vindicate the researcher's theory, but rather makes use of theory and observationally-gained knowledge to do something normatively desirable.

To recap, the second stage of a problem-solving research program involves trying to establish the real-world relevance of mechanisms observationally, with priority given to mechanisms that are potentially amenable to intervention. The skills one would need to develop for work in this stage include

- Using analytical methods to define potential mechanisms that would be potentially amenable to intervention,

- Representing mechanisms as DAGs and deriving observable implications and hypotheses that can be used to test the relevance of the mechanism, whether in the form of causal relationships or induced, non-causal relationships,
- Mastering observational identification strategies and evaluating research designs for their ability to identify causal relationships that a DAG represents, including attending to subtle issues of effective samples, local effects, commensurability problems, compensating differentials, etc., and
- Mastering observational research designs for capturing induced, potentially non-causal relationships that a DAG implies.

While most political science methodology focuses on this stage of research, the problem-solving approach views this stage as means toward defining potential interventions that could help to address the problem. That being the case, the goal is to evaluate the plausibility of a path toward addressing a problem, rather than trying to nail down causal processes definitively. Spirling and Stewart (2022) associate these tasks with what philosophers of science refer to as "inference to the best explanation," which refers to using evidence to guide in the determination of whether one or another account (which could involve multiple causal mechanisms) is most compelling as an explanation. Observational analyses will always involve ambiguities regarding causal identification, and political scientists should learn about partial identification strategies for inference under ambiguity (Manski, 2013; Duarte, Finkelstein, Knox, Mummolo, and Shpitser, 2021).

5 Testing Interventions

The descriptive and observational work described above in stages one and two lead to the design of interventions intended to mitigate the motivating problem. The end goal is to establish candidate interventions that can help to mitigate the problem. The proof in the pudding ultimately comes in the intervening—as Hacking (1983) explains, knowledge is secured when it is usable for generating desired outcomes.

Opportunities to test the proposed interventions may be found through collaboration with governmental agencies or non-governmental organizations. When such opportunities are found, we move to stage three of a problem-solving research program. The goal here is to test whether a proposed intervention can achieve at least a minimum target of improvement with respect to the motivating problem. Passing such a test would warrant having the intervention either tested or possibly implemented at larger scale. The empirical work here is *interventionist-causal*. Randomized controlled trials (RCTs) are an ideal approach, although designed regression discontinuity studies or designed difference-in-differences or synthetic control studies are second-best options when constraints prevent randomization.

The basic steps of stage three involve synthesizing what has been learned observationally and translating it into a test of an intervention concept. One uses theory and data to argue for the importance of a particular mechanism and opportunity to intervention, and then, on the basis of that mechanism, to define an intervention strategy. To test effectiveness, one defines primary and secondary outcomes of interest, ideally outcomes that are similar to those used in the descriptive phase of the research program to characterize the nature and extent of the problem. Then, one defines causal effects of interest, possibly including conditional or mediated effects, a randomization (or, if impossible, an observational design) strategy for identifying these effects, and then an estimation and testing strategy for assessing these effects. One can define a minimal effect size that would justify the costs of intervention and establish that as a benchmark against which to test. The quality of one's design is a depends on its unbiasedness and statistical power for detecting the minimal effect size of interest. The test informs a decision to proceed with testing or possibly implementation at a larger scale.

In a manner similar to normative analysis of problems that motivate research programs, an important skill set is in vetting the proposed intervention-based study in ethical terms. I agree with McDermott and Hatemi (2020) and APSA Ad Hoc Committee on Human Subjects Research (2020) that it is crucial to have ethical standards for field-interventionist research, but it should be clear that the problem-solving approach does not relegate political science research to seeking merely "to understand, not change, public outcomes" (McDermott and Hatemi, 2020, p. 30020). As argued above, testing interventions that are designed to bring about real-world change most certainly is, and from the problem-solving perspective, should certainly be central to what political scientists do. Problem-solving researchers can take guidance from colleagues in economics, including Asiedu, Karlan, Lambon-Quayefio, and Udry (2021), who provide guidance on ethical standards reporting that takes for granted that researchers are interested to test interventions that might matter in the real world.

Another important skill area is in managing collaborations with practitioners in governmental agencies or non-governmental organizations. Such collaborations are how political scientists can combine their theory and methods expertise with practitioners' implementation expertise and their connections to stakeholders whose lives would be affected by the intervention. Humphreys (2015) offers a systematic analysis of ethical responsibilities across such collaborations. The symposium edited by Davis and Michelitch (2022) offers a set of important reflections on how political scientists should take their positionality into account relative to those whose lives are affected by their research and their research partners.

The key methodological skill set comprises methods for design-based inference of RCTs, with foundational reference texts including Duflo, Glennerster, and Kremer (2007), Gerber and Green (2012), Imbens and Rubin (2015), and Athey and Imbens (2017)). Ethically speaking, interventionist research should proceed cautiously, and so a critical area for methodological development is in ways to learn from small-scale and short-term trials. This emphasis runs somewhat counter to current trends in developing methods for larger and larger datasets, although new methods for studying small trials do sometimes take advantage of machine learning methods. Novel approaches to studying small or short-term RCTs include methods for designing proxies for long term outcomes using measures that pick up on short-term variation (Athey et al., 2019), improving power through better stratification and covariate adjustment (Bloniarz, Liu, Zhang, Sekhon, and Yu, 2016; Harshaw, Sävje, Spielman, and Zhang, 2019), constructing tests that leverage multiple outcomes (Caughey, Dafoe, and Seawright, 2017; Casey, Glennerster, and Miguel, 2012), and combining observational and RCT data to boost overall power or power for group-specific effects (Rosenman, Basse, Owen, and Baiocchi, 2020; Asimovic, Ditlmann, and Samii, 2022). The relevant standard for power is the *minimal effect size of interest*, which is defined on the basis of what stakeholders consider to be an effect size that would warrant introducing an intervention at scale. A complicating factor is that, often, what are most relevant are conditional effects, which have strong power requirements. A related skill area is is in interpreting the implications of results from feasible, relatively small scale RCTs for what might happen at scale, which may introduce questions of equilibrium adjustment. This is similar to the work in macroeconomics and applied microeconomics on combining micro-level causal evidence with theoretical models to inform predictions about macro level policies (Nakamura and Steinsson, 2018) or long-run effects of policies (Wolpin, 2013, Sec 3.3).

Not all aspects of every problem lend themselves to precisely the same methodological strategies. For example the design of civil war peace settlements, which is squarely a political science topic, cannot be directly tested with conventional RCTs. In such cases, an interventionist-causal learning agenda would have to combine indirect evidence, for example from lab experimentation, with more elaborate theory and especially careful attention to details of individual cases. That said, there are important component or complementary problems that could proceed in more conventional ways, for example the study of community-level or individual-level post-conflict reintegration intervention (Blattman, 2022).

To recap, stage three of a problem-solving research program involves designing and executing tests of interventions. The skills needed here include allow researchers to do the following rigorously:

- Motivating an intervention concept based on evidence on mechanisms, ethics, and opportunity,
- Establishing relevant partnerships with stakeholder practitioners, and
- Designing a test, ideally an RCT, that is unbiased and well-powered for a minimal effect size of interest and that offers insights on impact at scale.

6 Conclusion

The aim here was to use a "problem-solving" framework for assessing the practical and normative value of political science research programs and for drawing connections between different areas of empirical methodology. Research programs can be organized around real-world problems, normatively defined. Within a research program, *descriptive* research motivates attention to the problem, *observational-causal* research helps to establish relevant mechanisms perpetuating the problem, and *intervention-causal* research tests interventions to try to mitigate the problem.

Rather than viewing each of these areas of methodology as somehow competing or at odds with each other, the problem-solving approach views them each as essential. Normative theory plays a role in motivating problems, and positive theory plays a role in making sense of empirical results and proposing generalizable mechanisms. The problem-solving approach proposes that political scientists can make societal contributions on the basis of their expertise in theory and empirical methods.

Political scientists frequently argue about whether a given study is "useful." Is a cross-national regression study that uses aggregate indices and no clear source of exogenous variation useful? Is a field experiment studying an intervention to affect political participation a small country useful? Examples like these are presented as ways to judge approaches to empirical research in a categorical sense. The challenge of course is to have some way to assess "useful" that is not merely personal taste. If political scientists accept the responsibility of contributing to addressing societal problems, then we can start to put some priority on questions. Useful research is that which moves us in the direction of defining interventions that can help to remedy societal problems. The form and content of such research can vary considerably, even within a single problem-focused research program.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Adcock, R. and D. Collier (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American political science review 95*(3), 529–546.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

APSA Ad Hoc Committee on Human Subjects Research (2020). *Principles and Guidance for Human Subjects Research*.

Aronow, P. M. and C. Samii (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science 60*(1), 250–267.

Ashworth, S., C. R. Berry, and E. B. de Mesquita (2021). *Theory and Credibility: Integrating Theoretical and Empirical Social Science*. Princeton University Press.

Asiedu, E., D. Karlan, M. Lambon-Quayefio, and C. Udry (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences 118*(29), e2024570118.

Asimovic, N., R. Ditlmann, and C. Samii (2022). Estimating the effect of intergroup contact over years: Evidence from a youth program in israel.

Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2019). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, National Bureau of Economic Research.

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, Volume 1, pp. 73–140. Elsevier.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto (2021). Conjoint survey experiments. *Advances in experimental political science 19*.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of political economy 101*(3), 385–409.

Bisbee, J. and J. M. Larson (2017). Testing social science network theories with online network data: An evaluation of external validity. *American political science review 111*(3), 502–521.

Blair, G., K. Imai, and Y.-Y. Zhou (2015). Design and analysis of the randomized response technique. *Journal of the American Statistical Association 110*(511), 1304–1319.

Blattman, C. (2022). *Why We Fight: The Roots of War and the Paths to Peace*. Penguin.

Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390.

Cantoni, E. and V. Pons (2021). Strict id laws don't stop voters: Evidence from a us nationwide panel, 2008–2018. *The Quarterly Journal of Economics 136*(4), 2615–2660.

Casey, K., R. Glennerster, and E. Miguel (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics 127*(4), 1755–1812.

Caughey, D., A. Dafoe, and J. Seawright (2017). Nonparametric combination (npc): A framework for testing elaborate theories. *The Journal of Politics 79*(2), 688–701.

Cohen, E. (2021). Measuring congressional gender bias from candidate to representative. Technical report, Emory University.

Cowen, T. (1998). Do economists use social mechanisms to explain? *Social mechanisms: An analytical approach to social theory*, 125–146.

Davis, J. and K. Michelitch (2022). Introduction to field experiments: Thinking through identity and positionality. *PS: Political Science & Politics*, 1–6.

De Mast, J. and J. Lokkerbol (2012). An analysis of the six sigma dmaic method from the perspective of problem solving. *International Journal of Production Economics 139*(2), 604–614.

De Mesquita, E. B. and S. A. Tyson (2020). The commensurability problem: Conceptual difficulties in estimating the effect of behavior on behavior. *American Political Science Review 114*(2), 375–391.

Delaney, J. and P. J. Devereux (2022). Rank effects in education: What do we know so far?

Duarte, G., N. Finkelstein, D. Knox, J. Mummolo, and I. Shpitser (2021). An automated approach to causal inference in discrete settings. *arXiv preprint arXiv:2109.13471*.

Duflo, E. (2017). The economist as plumber. *American Economic Review 107*(5), 1–26.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics 4*, 3895–3962.

Elwert, F. and C. Winship (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology 40*, 31.

Enamorado, T., B. Fifield, and K. Imai (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review 113*(2), 353–371.

Gelman, A. (2010). A data visualization manifesto. *Statistical Modeling, Causal Inference, and Social Science https://statmodeling.stat.columbia.edu/2010/05/31/a_data_visualiz/.*

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation.* WW Norton.

Golden, M. A. (2001). Why do trade unions call strikes. *Political science as puzzle solving*, 43.

Greenland, S. and J. Pearl (2014). Causal diagrams. *Wiley StatsRef: Statistics Reference Online*, 1–10.

Grofman, B. (2001). *Political Science as Puzzle Solving.* University of Michigan Press.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science.* Cambridge university press.

Harshaw, C., F. Sävje, D. Spielman, and P. Zhang (2019). Balancing covariates in randomized experiments with the gram–schmidt walk design. *arXiv preprint arXiv:1911.03071*.

Healy, K. (2018). *Data visualization: a practical introduction.* Princeton University Press.

Ho, D. E., K. M. Quinn, et al. (2008). Measuring explicit political positions of media. *Quarterly Journal of Political Science 3*(4), 353–377.

Humphreys, M. (2015). Reflections on the ethics of social experimentation. *Journal of Globalization and Development 6*(1), 87–112.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

King, G., C. J. Murray, J. A. Salomon, and A. Tandon (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American political science review 98*(1), 191–207.

Knox, D., W. Lowe, and J. Mummolo (2020). Administrative records mask racially biased policing. *American Political Science Review 114*(3), 619–637.

Kymlicka, W. (2002). *Contemporary political philosophy: An introduction.* oxford: oxford University Press.

Laver, M. (2014). Measuring policy positions in political space. *Annual Review of Political Science 17*, 207–223.

Mahdavi, P. (2019). Scraping public co-occurrences for statistical network analysis of political elites. *Political Science Research and Methods 7*(2), 385–392.

Manski, C. F. (2013). *Public policy in an uncertain world: analysis and decisions.* Harvard University Press.

Marshall, J. (2022). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science*.

McDermott, R. and P. K. Hatemi (2020). Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms. *Proceedings of the National Academy of Sciences 117*(48), 30014–30021.

Morgan, S. L. and C. Winship (2015). *Counterfactuals and causal inference.* Cambridge University Press.

Moynihan, D. P. (2022). How can scholars help to embed institutions of public-sector change?(or things i wish i'd known when i was a grad student). *Perspectives on Public Management and Governance.*

Nakamura, E. and J. Steinsson (2018). Identification in macroeconomics. *Journal of Economic Perspectives 32*(3), 59–86.

Pearl, J. and D. Mackenzie (2018). *The book of why: the new science of cause and effect.* Basic books.

Pettigrew, S. (2017). The racial gap in wait times: why minority precincts are underserved by local election officials. *Political Science Quarterly 132*(3), 527–547.

Rodrigues, D., N. Kreif, A. Lawrence-Jones, M. Barahona, and E. Mayer (2022). Reflection on modern methods: constructing directed acyclic graphs (dags) with domain experts for health services research. *International Journal of Epidemiology 51*(4), 1339–1348.

Rosenbaum, P. R. (2010). *Design of observational studies*, Volume 10. Springer.

Rosenman, E., G. Basse, A. Owen, and M. Baiocchi (2020). Combining observational and experimental datasets using shrinkage estimators. *arXiv preprint arXiv:2002.06708*.

Samii, C. (2016). Causal empiricism in quantitative research. *The Journal of Politics 78*(3), 941–955.

Schedler, A. (2012). Judgment and measurement in political science. *Perspectives on politics 10*(1), 21–36.

Simon, H. A. (2019). *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird.* MIT press.

Sniderman, P. M. (2018). Some advances in the design of survey experiments. *Annual Review of Political Science 21*, 259–275.

Spirling, A. and B. M. Stewart (2022). What good is a regression? inference to the best explanation and the practice of political science research.

Weisberg, H. F. (2009). *The total survey error approach: A guide to the new science of survey research*. University of Chicago Press.

Westwood, S. J., J. Grimmer, M. Tyler, and C. Nall (2022). Current research overstates american support for political violence. *Proceedings of the National Academy of Sciences 119*(12), e2116870119.

Wolpin, K. I. (2013). *The limits of inference without theory*. MIT Press.

Yau, N. (2011). *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons.

Zeller, R. A. and E. G. Carmines (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge University Press.