Oxford Handbooks Online

Abstract and Keywords

Political methodology offers techniques for clarifying the theoretical meaning of concepts such as revolution and for developing definitions of revolutions. It also provides descriptive indicators for comparing the scope of revolutionary change, and sample surveys for gauging the support for revolutions. It then presents an array of methods for making causal inferences that provide insights into the causes and consequences of revolutions. An overview of the book is given. Topics addressed include social theory and approaches to social science methodology; concepts and development measurement; causality and explanation in social research; experiments, quasi-experiments, and natural experiments; general methods of quantitative tools for causal and descriptive inference; quantitative tools for causal and descriptive inference; qualitative tools for causal inference; and organizations, institutions, and movements in the field of methodology. In general, the *Handbook* provides overviews of specific methodologies, but it also emphasizes three things: utility for understanding politics, pluralism of approaches, and cutting across boundaries. This volume discusses interpretive and constructivist methods, along with broader issues of situating alternative analytic tools in relation to an understanding of culture.

Keywords: political science, political methodology, causal inference, descriptive inference, social theory, revolution
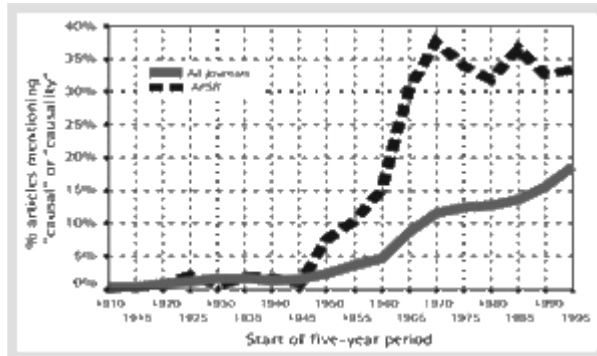
"You say you want a revolution

Well, you know,

We all want to change the world."

The Beatles

date: 02 January 2017

PEOPLE of the 1960s generation did not argue much with the Beatles—we listened to them with rapt attention. But are they right? What do they mean by a revolution? Do we all want to change the world? What would change the world? Would the result be good or bad?

Click to view larger

Fig. 1.1. Growth of "causal thinking" in JSTOR articles from 1910 to 1999

Political methodology provides the practicing political scientist with tools for attacking all these questions, although it leaves to normative political theory the question of what is ultimately good or bad. Methodology provides techniques for clarifying the theoretical meaning of concepts such as revolution and for developing definitions of revolutions. It offers descriptive indicators for comparing the scope of revolutionary change, and sample surveys for gauging the support for revolutions. And it offers an array of methods for making causal inferences that provide insights into the causes and consequences of revolutions. All these tasks are important and strongly (p. 4) interconnected. While causal inference is fundamental in political science, making good inferences depends entirely on adequate conceptualization and measurement of the phenomena under study—tasks that receive substantial attention in this volume. Yet time and again our authors return to the question of what might constitute a valid causal inference using qualitative or quantitative data, small-*N* or large-*n* data, in-depth interviews or sample surveys, historical narratives or experimental data.

Although not all of modern political science is about causal inference, between 1995 and 1999 about 33 percent of the articles in the *American Political Science Review (APSR)* mentioned these words and 19 percent of all the journal articles in JSTOR for this period mentioned them. The proportions rise to 60 percent for all journals and 67 percent for the *APSR* if we add the words "cause" or "causes," but these words do not have the same technical meaning as "causal" or "causality" so we will stick with the narrower measure of our concept, even though it might be an underestimate of the scope of causal thinking.[1] As shown in Figure 1.1, the concern with causality is increasing, and the mentions of these terms grew rapidly from less than 2 percent of JSTOR articles from 1910 to 1950 to an increasing proportion from 1950 onwards, with the *APSR* apparently leading the way.

What is behind this dramatic increase in mentions of "causal" or "causality?" Does it tell us something meaningful about political science in the twentieth century? Have we measured a useful concept (i.e. "causal thinking in political science") (p. 5) with our JSTOR search? Have we described accurately the rise of causal thinking in the twentieth century? Can we explain this rise? The methods contained in this handbook are expressly designed to answer social science questions of this sort. Our discussion of causality may be just a "toy example," but it does have the virtue that it is familiar to and perhaps interesting to political scientists. And it has the additional virtue that explaining the increasing concern with a new perspective such as "causal thinking" within political science is a miniature and simpler version of explaining the rise of "revolutionary" perspectives in politics—the emergence of eighteenth-century liberalism, nineteenth-century socialism, early to mid-twentieth- century New Deal liberalism, late twentieth-century neoliberalism, and the modern environmental movement. If we can understand the difficulties of *explaining* the rise of causal thinking within political science, indeed the difficulties of merely *describing* whether or not causal thinking actually increased during the twentieth century, we will not only provide an overview of this handbook, but we will also learn a lot about what methodologists can contribute to doing political science research. If along the way the reader grimaces over some of our methodological approaches, we hope this reaction has the effect of raising questions about what can and cannot be done with these methods. Perhaps it will also help us all develop some modesty about what our craft can accomplish.

1 Social Theory and Approaches to Social Science Methodology

How do we think about explaining the rise of causal thinking in political science? One place to start is with social theory which asks questions about the ontology and epistemology of our enterprise. Ontology deals with the things that we think exist in the world, and epistemology with how we come to know about those things. Hardin (Chapter 2) suggests that we should start social science inquiry with individuals, their motivations, and the kinds of transactions they undertake with one another. He starts with self-interest (although he quickly suggests that there are other motivations as well), and this provides a useful starting place for understanding the increasing focus on causality in political science. Self-interest suggests that people publish in order to advance their careers and that they will do what is necessary to achieve that end, but it begs the question of why causal thinking is a common goal of the political science profession.

date: 02 January 2017

Hardin describes four basic schools of social theory: conflict, shared-values, exchange, and coordination theories. Several of these might help to explain why political scientists have adopted causal thinking as a goal for their enterprise. Political scientists might be adhering to shared "scientific" values about understanding the (p. 6) world through the exploration of causal relationships. And this scientific value might have become important to science in the twentieth century because it allowed humans to manipulate their world and to shape it in their self-interest. According to this explanation, the social sciences simply adopted this value because, well, it was valuable. Alternatively, political scientists might be exchanging their causal knowledge for resources garnered from the larger society. Or they might be coordinating on the topic of causality in order to have a common standard for evaluating research, although this leaves open why they chose this solution to the coordination problem. One answer might be that a focal point was created through the invention of some convenient tool that promised to help political scientists with their research. Two obvious methodological tools of the early twentieth century are correlation analysis (Pearson 1909) and regression analysis (Pearson 1896; Yule 1907), although as we shall see, only regression analysis provided at least rhetorical support for causal inference.

Bevir (Chapter 3) provides some explanations for the rise of causal thinking as the "behavioral revolution's" reaction to the nineteenth century's teleological narratives about history ("developmental historicism") and early twentieth-century emphasis on classifications, correlations, and systems ("modernist empiricism"). The behavioral revolution took a somewhat different direction and emphasized general theories and the testing of causal hypotheses. Bevir's chapter suggests that the rise of causal thinking might have been a corollary of this development. But Bevir warns that there are new currents in philosophy which have moved beyond behavioralism.

De Marchi and Page (Chapter 4) explore one kind of mathematical modeling, agent-based modeling, that has become increasingly common in political science. We might have included chapters on other theoretical perspectives (rational choice, social network modeling, historical narratives, etc.) but this one was especially apt for a methodology handbook since agent-based modeling is not only an approach to modeling; it is also a way of simulating models to generate testable hypotheses and even of generating data that can then be analyzed. Agent-based models suggest that we should think of political scientists as agents with goals who interact according to some rules—including rule-changing rules. These "rule-changing rules" might include changes in what is valued or in how people coordinate—such as a change towards emphasizing causal thinking over other kinds of inquiry.

date: 02 January 2017
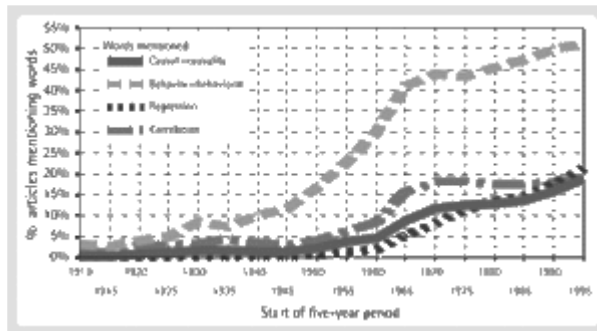
Click to view larger

Fig. 1.2. Growth of mentions of words related to causal thinking in political science in JSTOR articles from 1910 to 1999

Three possible causes of the increased emphasis on causality follow from this discussion. Two of them look to the development of a new tool, either regression or correlation, that made it easier to determine causality so that more scholars focused upon that problem. The third suggests value change with the rise of behavioralism. There may be other plausible explanations, but these provide us with a trio of possibilities for developing our example. Indeed, these categories of explanation— new inventions and new values—crop up again and again in social science. The rise of capitalism, for example, has been explained as the result of inventions such as markets, corporations, and industrial processes that made individual accumulation possible, and it has been explained as the result of an emphasis on particular values such as a protestant ethic that valued accumulation and accomplishment. (p. 7)

2 Concepts and Measurement

To proceed with our investigation of the rise in causal thinking, we must clarify our concepts and develop measures. Our concepts are "the study of causality in political science," the use of the tools of "regression analysis" or "correlation," and changes in values due to the "behavioral revolution." Continuing with what we have already done, we measure them using word searches in JSTOR. For regression and correlation, we look for "regression" or "correlation."[2] We risk, of course, the possibility that these terms are being used in nonstatistical ways ("regression to his childhood" or "the correlation of forces"), but we assume for the moment that these uses stay relatively constant over time.

In order to determine the utility of this approach, we focus on the definition of "behavioral revolution," but if we had more space we could have added similar discussions about measuring "the study of causality in political science" or "correlation" or "regression." To measure the extent of the behavioral revolution, we look for the

date: 02 January 2017

words "behavior" or "behavioral." When we count these various terms over the ninety years we get the results in Figure 1.2.

Goertz (Chapter 5) provides guidance on how to think about our concepts. Not surprisingly, he tells us that we must start by thinking about the theory embedded in the concept, and we must think about the plausible method for aggregating indicators (p. 8) of the concept. For measuring the extent of the "the behavioral revolution" we want to measure those habits and perspectives of inquiry that distinguished those researchers who were concerned with general theories of behavior from those who went before them. Simply counting words may seem like a poor way to do this—at first blush it would seem that we should use a more sophisticated method that codes articles based on whether or not they proposed general hypotheses, collected data to test them, and carried out some tests to do just that. At the very least, we might look for the word "behavioralism" or "behaviorism" to make sure that the authors subscribed to the movement. But from 1910 to 1999, "behavioralism" is only mentioned in 338 articles— out of a total of 78,046 (about 0.4 percent). And "behaviorism" is mentioned in even fewer articles, which is not too surprising since political scientists (although not psychologists) tended to prefer the term "behavioralism."

The words "behavioral" and "behavior" turn out to be better measures as judged by tests of criterion and convergent validity (Jackman, Chapter 6). The word behavioral is mentioned in 8.9 percent of the articles and the word "behavior" in 31.3 percent. These two words (behavior and behavioral) are almost always mentioned when the criterion words of behavioralism or behaviorism are mentioned (about 95 percent of the time). Moreover, in a test of convergent validity, the articles of those people known to be leaders of the behavioral movement used these terms more frequently than the authors of the average article. Between 1930 and 1990, we find that the average article mentioned one or more of the four terms 33 percent of the time, but the articles by the known behavioralists mentioned one or more of the four terms 66 percent of the time.[3] Hence, these words appear to be closely related to the behavioral movement, and we will often refer to mentions of them as indicators of "behavioralism." Similarly, we will often refer to mentions of "causal" and "causality" as indicators of "causal thinking."

In our running example, we used JSTOR categories to describe scientific disciplines (political science, sociology, etc.) and to classify journals and items (as articles or book reviews or editorials) according to these categories. Collier, LaPorte, and Seawright (Chapter 7) and Ragin (Chapter 8) remind us that these are important decisions with significant implications for conceptualization and calibration.

Collier, LaPorte, and Seawright (Chapter 7) discuss categories and typologies as an optic for looking at concept formation and measurement. Working with typologies is crucial not

only to the creation and refinement of concepts, but it also contributes to constructing categorical variables involving nominal, partially ordered, and ordinal scales. Although typologies might be seen as part of the qualitative tradition of research, in fact they are also employed by quantitative analysts, and this chapter (p. 9) therefore provides one of the many bridges between these two approaches that are crucial to the approach of this handbook.

Ragin (Chapter 8) distinguishes between "measurement" and "calibration," arguing that with calibration the researcher achieves a tighter integration of measurement and theory. For example, a political science theory about "developed countries" will probably not be the same as a theory about "developing countries," so that careful thought must be given to how the corresponding categories are conceptualized, and how countries are assigned to them. In our running example, articles in political science will probably be different from those in other disciplines, so care must be taken in defining the scope of the discipline. Yet we have rather cavalierly allowed JSTOR to define this membership, even though by JSTOR's categorization, political science thereby includes the journals *Social Science History*, the *Journal of Comparative Law*, and *Asian Studies*. We have also allowed JSTOR to treat articles as examples of "causal thinking" when they have at least one mention of "causal" or "causality" even though there might be a substantial difference between articles that mention these terms only once versus those that mention them many times. Alternative calibration decisions are certainly possible. Perhaps only journals with political, politics, or some similar word in their titles should be considered political science journals. Perhaps we should require a threshold number of mentions of "causal" or "causality" before considering an article as an example of "causal thinking." Perhaps we should revisit the question of whether "cause" and "causes" should be used as measures of "causal thinking." Ragin provides a "fuzzy-set" framework for thinking about these decisions, and thereby offers both direct and indirect methods for calibration.

Jackman (Chapter 6) also focuses on measurement, starting from the classic test theory model in which an indicator is equal to a latent variable plus some error. He reminds us that good measures must be both valid and reliable, and defines these standards carefully. He demonstrates the dangers of unreliability, and discusses the estimation of various measurement models using Bayesian methods. Jackman's argument reminds us to consider the degree to which our counts of articles that mention specific words represent the underlying concepts, and he presents a picture of measurement in which multiple indicators are combined—typically additively— to get better measures of underlying concepts. Goertz's chapter suggests that there is an alternative approach in which indicators are combined according to some logical formula. Our approach to measuring behavioralism at the article level has more in common with Goertz's approach because it requires that either "behavior" or "behavioral" be present in the article, but it has more

date: 02 January 2017

in common with Jackman's approach when we assume that our time series of proportions of articles mentioning these terms is a valid and relatively reliable measure of the degree to which behavioralism has infused the discipline.

Poole (Chapter 9), Jackman (Chapter 6), and Bollen et al. (Chapter 18) consider whether concepts are multidimensional. Any measurement effort should consider this possibility, but political scientists must be especially careful because the dimensionality of politics matters a great deal for understanding political contestation. To take just one example, multidimensional voting spaces typically lead to voting cycles (p. 10) (Arrow 1963) and "chaos" theorems (McKelvey 1979; Schofield 1983; Saari 1999) for voting systems. Poole reviews how the measurement of political issue spaces has developed in the past eighty years through borrowings from psychometrics (scaling, factor analysis, and unfolding), additions from political science theories (the spatial theory of voting and ideal points), and confronting the special problems of political survey, roll-call, and interest-group ratings data. Bollen et al. show how factor analysis methods for determining dimensionality can be combined with structural equation modeling (SEM).

Table 1.1. Two dimensions of political science discourse, 1970–1999		
	Component	
	Causal	**Interpretive**
Narrative	.018	.759
Interpretive	.103	.738
Causal/causality	.700	.105
Hypothesis	.750	−.073
Explanation	.701	.131

Extraction method: principal component analysis. Rotation method: oblimin with Kaiser normalization.

There does not seem to be any obvious need to consider dimensions for our data, but suppose we broaden our inquiry by asking whether there are different dimensions of political science discourse. Based upon our relatively detailed qualitative knowledge of "political science in America," we chose to search for all articles from 1970 to 1999 on

date: 02 January 2017

five words that we suspected might have a two-dimensional structure: the words "narrative," "interpretive," "causal or causality," "hypothesis," and "explanation." After obtaining their correlations across articles,[4] we used principal components and an oblimin rotation as described in Jackman (Chapter 6). We found two eigenvalues with values larger than one which suggested a two dimensional principal components solution reported in Table 1.1. There is a "causal dimension" which applies to roughly one-third of the articles and an "interpretive" dimension which applies to about 6 percent of the articles.[5] Although we expected this two- dimensional structure, we were somewhat surprised to find that the word "explanation" was almost entirely connected with "causal or cauality" and with "hypothesis." And we were surprised that the two dimensions were completely distinctive, since they are essentially uncorrelated at .077. Moreover, in a separate analysis, we found (p. 11) that whereas the increase in "causal thinking" occurred around 1960 or maybe even 1950 in political science (see Figure 1.1), the rise in the use of the terms "narrative" and "interpretive" came in 1980.[6]

This result reminds us that "causal thinking" is not the only approach to political science discourse. Our volume recognizes this by including chapters that consider historical narrative (Mahoney and Terrie, Chapter 32) and intensive interviewing (Rathbun, Chapter 28), but there is also a rich set of chapters in a companion volume, the *Oxford Handbook of Contextual Political Analysis*, which the interested reader might want to consult.

This discussion leads us to think a bit more about our measure of "causal thinking." The chapters on "Concepts and Measurement" suggest that we have been a bit cavalier in our definition of concepts. Perhaps we should be thinking about measuring "scientific thinking" instead of just "causal thinking." How can we do that? In fact, like many researchers, we started with an interesting empirical fact (i.e. the mentions of "causal" and "causality" in political science articles), and worked from there. At this point, some hard thinking and research (which will be mostly qualitative) about our concepts would be useful. Philosophical works about the nature of science and social science should be consulted. Some well-known exemplars of good social science research should be reviewed. And something like the following can be done.

Based upon our reflections about the nature of scientific thinking (and the factor analysis above), we believe that the words "hypothesis" and "explanation" as well as "causal or causality" might be thought of as indicators of a "scientific" frame of mind.[7] Consider the frequency of these words in all articles in JSTOR in various disciplines from 1990 to 1999. Table 1.2 sorts the results in terms of the discipline with the highest use of any of the words at the top of the table. Note that by these measures, ecology and evolutionary biology, sociology, and economics are most "scientific" while "history," "film studies," and "performing arts" are least "scientific." Also note that the highest figures in each row

date: 02 January 2017

(excluding the final column) are in bold. Note that we put "scientific" in quotations because we want to emphasize our special and limited definition of the term.

Ecology and evolutionary biology and economics refer to "hypothesis" to a greater degree than other disciplines which mention "explanation" more. But also note that political science (17.2 percent) and sociology (25.2 percent) tend to be high in mentions of "causal" or "causality." In contrast, "performing arts" has a 3.6 percent rate of mention of "causal" or "causality" and "film studies" has a 5.8 percent rate.

As researchers, we might at this point rethink our dependent variable, but we are going to stay with mentions of "causal or causality" for two reasons. First, these words (p. 12) come closest to measuring the concerns of many authors in our book. Second, the narrowness of this definition (in terms of the numbers of articles mentioning the terms) may make it easier to explain. But the foregoing analysis (including our discussion of narrative and interpretive methods) serves as a warning that we have a narrow definition of what political science is doing.

date: 02 January 2017

Table 1.2. Mentions of "scientific" terms in various disciplines, 1990–9 (%)				
	Causal/ causality	Hypothesis	Explanation	Any of these words
Ecology & evolutionary biology	7.9	**50.0**	31.7	62.3
Sociology	25.1	37.3	**40.6**	61.7
Economics	10.0	**44.6**	36.1	61.2
Political science	17.2	27.3	**38.0**	52.2
Anthropology	12.5	22.8	**39.1**	52.0
History	8.1	14.3	**36.4**	44.6
Film studies	5.8	5.3	**19.3**	25.4
Performing arts	3.6	4.8	**18.5**	23.8

Source: Searches of JSTOR archive by authors.

3 Causality and Explanation in Social Research

Brady (Chapter 10) presents an overview of causal thinking by characterizing four approaches to causal inference. The Humean regularity approach focuses on "lawlike" constant conjunction and temporal antecedence, and many statistical methods—pre-eminently regression analysis—are designed to provide just the kind of information to satisfy the requirements of the Humean model. Regression analysis can be used to determine whether a dependent variable is still correlated ("constantly conjoined") with an independent variable when other plausible causes of the dependent variable are held constant by being included in the regression; and time-series regressions can look for

date: 02 January 2017

temporal antecedence by regressing a dependent variable on lagged independent variables.

In our running example, if the invention of regression analysis actually led to the emphasis upon causality in political science, then we would expect to find two things. First in a regression of "causal thinking" (that is, mentions of "causal or causality") on mentions of "regression," mentions of "correlation," and mentions of "behavioralism," we expect to find a significant regression coefficient on the "regression" variable. Second, we would expect that the invention of the method of regression and its (p. 13) introduction into political science would pre-date the onset of "causal thinking" in political science. In addition, in a time-series regression of mentions of "causal thinking" on lagged values of mentions of "regression," "correlation," and "behavioralism" we would expect a significant coefficient on lagged "regression." We shall discuss this approach in detail later on.

The counterfactual approach to causation asks what would have happened had a putative cause not occurred in the most similar possible world without the cause. It requires either finding a similar situation in which the cause is not present or imagining what such a situation would be like. In our running example, if we want to determine whether or not the introduction of regression analysis led to an efflorescence of causal thinking in political science, we must imagine what would have happened if regression analysis had not been invented by Pearson and Yule. In this imagined world, we would not expect causal thinking to develop to such a great extent as in our present world. Or alternatively, we must find a "similar" world (such as the study of politics in some European country such as France) where regression was not introduced until much later than in the United States. In this most similar world, we would not expect to see mentions of "causal thinking" in the political science literature until much later as well.

The manipulation approach asks what happens when we actively manipulate the cause: Does it lead to the putative effect? In our running example, we might consider what happened when the teaching of regression was introduced into some scholarly venue. When graduate programs introduced regression analysis, do we find that their new Ph.Ds focused on causal issues in their dissertations? Does the manipulation of the curriculum by teaching regression analysis lead to "causal thinking?"

Finally, as we shall see below, the mechanism and capacities approach asks what detailed steps lead from the cause to the effect. In our running example, it asks about the exact steps that could lead from the introduction of regression analysis in a discipline to a concern with causality.

date: 02 January 2017

Brady also discusses the INUS model which considers the complexity of causal factors. This model gets beyond simple necessary or sufficient conditions for an effect by arguing that often there are different sufficient pathways (but no pathway is strictly necessary) to causation—each pathway consisting of an *insufficient* but *nonredundant* part of an *unnecessary* but *sufficient* (INUS) condition for the effect.

Sekhon (Chapter 11) provides a detailed discussion of the Neyman—Rubin model of causal inference that combines counterfactual thinking with the requirement for manipulation in the design of experiments. This model also makes the basic test of a causal relationship a probabilistic one: whether or not the probability of the effect goes up when the cause is present.[8] Sekhon shows how with relatively weak assumptions (but see below) this approach can lead to valid causal inferences. He (p. 14) also discusses under what conditions "matching" approaches can lead to valid inferences, and what kinds of compromises sometimes have to be made with respect to generalizability (external validity) to ensure valid causal inferences (internal validity).

Freedman (Chapter 12) argues that "substantial progress also derives from informal reasoning and qualitative insights." Although he has written extensively on the Neyman—Rubin framework and believes that it should be employed whenever possible because it sets the gold standard for causal inferences, Freedman knows that in the real world, we must sometimes fall back on observational data. What do we do then? The analysis of large "observational" data-sets is one approach, but he suggests that another strategy relying upon "causal process observations" (CPOs) might be useful as a complement to them. CPOs rely on detailed observations of situations to look for hints and signs that one or another causal process might be at work. These case studies sometimes manipulate the putative cause, as in Jenner's vaccinations. Or they rule out alternative explanations, as in Semmelweis's rejection of "atmospheric, cosmic, telluric changes" as the causes for puerperal fever. They take advantage of case studies such as the death of Semmelweis's colleague by "cadaveric particles," Fleming's observation of an anomaly in a bacterial culture in his laboratory that led to the discovery of penicillin, or the death of a poor soul in London who next occupied the same room as a newly arrived and cholera infected seaman. Or a lady's death by cholera from what Snow considered the infected water from the "Broad Street Pump" even though she lived far from the pump but, it turned out, liked the taste of the water from the pump.

Hedström (Chapter 13) suggests that explanation requires understanding mechanisms which are the underlying "cogs and wheels" which connect the cause and the effect. The mechanism, for example, which explains how vaccinations work to provide immunity from an illness is the interaction between a weakened form of a virus and the body's immune system which confers long-term immunity. In social science, the rise in a candidate's popularity after an advertisement might be explained by a psychological process that

date: 02 January 2017

works on a cognitive or emotional level to process messages in the advertisement. Hedström inventories various definitions of "mechanism." He provides examples of how they might work, and he presents a framework for thinking about the mechanisms underlying individual actions.

In our running example, it would be useful to find out how regression might have become a tool for supposedly discovering causality. Some of the mechanisms include the following. Regression is inherently asymmetrical leading to an identification of the "dependent" variable with the effect and the "independent" variables with possible causes. The interpretation of regression coefficients to mean that a unit change in the independent variable would lead to a change in the dependent variable equal to the regression coefficient (everything else equal) strongly suggests that regression coefficients can be treated as causal effects, and it provides a simple and powerful way to describe and quantify the causal effect for someone. The names for regression techniques may have played a role from about 1966 onwards when there was a steady growth for the next twenty-five years in articles that described regression analyses as (p. 15) "causal models" or "causal modeling"[9] even though some authors would argue that the names were often seriously misleading—even amounting to a "con job" (Leamer 1983; Freedman 2005). And the relative ease with which regression could be taught and used (due to the advent of computers) might also explain why it was adopted by political scientists.

4 Experiments, Quasi-experiments, and Natural Experiments

Experiments are the gold standard for establishing causality. Combining R. A. Fisher's notion of randomized experiment (1925) with the Neyman—Rubin model (Neyman 1923; Rubin 1974; 1978; Holland 1986) provides a recipe for valid causal inference as long as several assumptions are met. At least one of these, the Stable Unit Treatment Value Assumption (SUTVA), is not trivial,[10] but some of the others are relatively innocuous so that when an experiment can be done, the burden of good inference is to properly implement the experiment. Morton and Williams (Chapter 14) note that the number of experiments has increased dramatically in political science in the last thirty-five years because of their power for making causal inferences.[11] At the same time, they directly confront the Achilles heel of experiments—their external validity. They argue that external validity can be achieved if a result can be replicated across a variety of data-sets and situations. In some cases this means trying experiments in the field, in surveys, or on the internet; but they also argue that the control possible in laboratory experimentation

date: 02 January 2017

can make it possible to induce a wider range of variation than in the field—thus increasing external validity. They link formal models with experimentation by showing how experiments can be designed to test them.

For Gerber and Green (Chapter 15) field experiments and natural experiments are a way to overcome the external validity limitations of laboratory experiments. They show that despite early skepticism about what could be done with experiments, social scientists are increasingly finding ways to experiment in areas such as criminal justice, the provision of social welfare, schooling, and even politics. But they (p. 16) admit that "there remain important domains of political science that lie beyond the reach of randomized experimentation." Gerber and Green review the Neyman—Rubin framework, discuss SUTVA, and contrast experimental and observational inference. They also discuss the problems of "noncompliance" and "attrition" in experiments. Noncompliance occurs when medical subjects do not take the medicines they are assigned or citizens do not get the phone calls that were supposed to to encourage their participation in politics. Attrition is a problem for experiments when people are more likely to be "lost" in one condition (typically, but not always, the control condition) than another. They end with a discussion of natural experiments where some naturally occurring process such as a lottery for the draft produces a randomized or nearly randomized treatment.

With the advice of these articles in hand, we can return to our running example. We are encouraged to think hard about how we might do an experiment to find out about the impact of new techniques (regression or correlation) or changes in values (the behavioral revolution) on causal thinking. We could, for example, randomly assign students to either a 1970s-style curriculum in which they learned about "causal modeling" methods such as regression analysis or a 1930s-style curriculum in which they did not. We could then observe what kinds of dissertations they produced. It would also be interesting to see which group got more jobs, although we suspect that human subjects committees (not to mention graduate students) would look askance at these scientific endeavors. Moreover, there is the great likelihood that SUTVA would be violated as the amount of communication across the two groups might depend on their assignment. All in all, it is hard to think of experiments that can be done in this area. This example reminds us that for some crucial research questions, experiments may be impossible or severely limited in their usefulness.

date: 02 January 2017

5 Quantitative Tools for Causal and Descriptive Inference: General Methods

Our discussion of the rise of causal thinking in political science makes use of the JSTOR database. Political science is increasingly using databases that are available on the internet. But scientific surveys provided political scientists with the first opportunities to collect micro-data on people's attitudes, beliefs, and behaviors, and surveys continue to be an immensely important method of data collection. Other handbooks provide information on some of these other methods of data collection, but the discussion of survey methods provides a template for thinking about data collection issues. Johnston (Chapter 16) considers three dimensions for data collection: mode, space, and time. For sample surveys, the modes include mail, telephone, in-person, (p. 17) and internet. Space and time involve the methods of data collection (clustered samples versus completely random samples) and the design of the survey (cross-sectional or panels). Beyond mode, space, and time, Johnston goes on to consider the problems of adequately representing persons by ensuring high response rates and measuring opinions validly and reliably through the design of high-quality questions.

In our running example, our data come from a computerized database of articles, but we could imagine getting very useful data from other modes such as surveys, in- depth interviews, or old college catalogs and reading lists for courses. Our JSTOR data provide a fairly wide cross-section of extant journals at different locations at any moment in time, and they provide over-time data extending back to when many journals began publishing. We can think of the data as a series of repeated cross-sections, or if we wish to consider a number of journals, as a panel with repeated observations on each journal. As for the quality of the data, we can ask, as does Johnston in the survey context about the veracity of question responses, whether our articles and coding methods faithfully represent people's beliefs and attitudes.

The rest of this section and all of the next section of the handbook discuss regression-like statistical methods and their extensions. These methods can be used for two quite different purposes that are sometimes seriously conflated and unfortunately confused. They can be used for *descriptive inferences* about phenomena, or they can be used to make *causal inferences* about them (King, Keohane, and Verba 1994). Establishing the Humean conditions of constant conjunction and temporal precedence with regression-like methods often takes pride of place when people use these methods, but they can also be thought of as ways to describe complex data-sets by estimating parameters that tell us important things about the data. For example, Autoregressive Integrated Moving Average

date: 02 January 2017

(ARIMA) models can quickly tell us a lot about a time series through the standard "p,d,q" parameters which are the order of the autoregression (p), the level of differencing (d) required for stationarity, and the order of the moving average component (q). And a graph of a hazard rate over time derived from an events history model reveals at a glance important facts about the ending of wars or the dissolution of coalition governments. Descriptive inference is often underrated in the social sciences (although survey methodologists proudly focus on this problem), but more worrisome is the tendency for social scientists to mistake description using a statistical technique for valid causal inferences. For example, most regression analyses in the social sciences are probably useful descriptions of the relationships among various variables, but they often cannot properly be used for causal inferences because they omit variables, fail to deal with selection bias and endogeneity, and lack theoretical grounding.

Let us illustrate this with our running example. The classic regression approach to causality suggests estimating a simple regression equation such as the following for cross-sectional data on all political science articles in JSTOR between 1970 and 1979. For each article we score a mention of either "causality or causal" as a one and no mention of these terms as a zero. We then regress these zero—one values of the "dependent variable" on zero—one values for "independent variables" measuring (p. 18) whether or not the article mentioned "regression," "correlation," or "behavioralism." When we do this, we get the results in column one in Table 1.3.

date: 02 January 2017

Table 1.3. Results of regressing whether "causal thinking" was mentioned among potential explanatory factors for 1970–1979—all political science journal articles in JSTOR

Independent variables	Regression coefficient (standard error)	
	One	**Two**
Behavior	.122 (.006)**	.110 (.006)**
Regression	.169 (.010)**	.061 (.021)*
Correlation	.157 (.008)**	.150 (.015)**
Behavior × regression		.135 (.022)**
Behavior × correlation		.004 (.017)
Regression × correlation		.027 (.021)
Constant	.022 (.008)**	.028 (.004)**
R^2/N	.149/ 12,305	.152/ 12,305

(***) Significant at .001 level;

(**) Significant at .01 level;

(*) Significant at .05 level.

If we use the causal interpretation of regression analysis to interpret these results, we might conclude that all three factors led to the emphasis on "causal thinking" in political science because each coefficient is substantively large and statistically highly significant. But this interpretation ignores a multitude of problems.

Given the INUS model of causation which emphasizes the complexity of necessary and sufficient conditions, we might suspect that there is some interaction among these variables so we should include interactions between each pair of variables. These interactions require that both concepts be present in the article so that a "regression × correlation" interaction requires that both regression and correlation are mentioned. The results from estimating this model are in column two of the table. Interestingly, only the

date: 02 January 2017

"behavior × regression" interaction is significant, suggesting that it is the combination of the behavioral revolution and the development of regression analysis that "explains" the prevalence of causal thinking in political science. (The three-way interaction is not reported and is insignificant.) Descriptively this result is certainly correct—it appears that a mention of behavioralism alone increases the probability of "causal thinking" in an article by about 11 percent, the mention of regression increases the probability by about 6 percent, the mention of correlation increases the probability by about 15 percent, and the mention of both behavioralism and regression together further increases the probability of causal thinking by about 13.5 percent.

But are these causal effects? This analysis is immediately open to the standard criticisms of the regression approach when it is used to infer causation: Maybe some other factor (or factors) causes these measures (especially "behavioral," "regression," and "causality") to cohere during this period. Maybe these are all spurious relationships which appear to be significant because the true cause is omitted from the equation. (p. 19) Or maybe causality goes both ways and all these variables are endogenous. Perhaps "causal thinking" causes mentions of the words "behavioral or behavior" and "regression" and "correlation."

Although the problem of spurious relationships challenged the regression approach from the very beginning (see Yule 1907), many people (including Yule) thought that it could be overcome by simply adding enough variables to cover all potential causes. The endogeneity problem posed a greater challenge which only became apparent to political scientists in the 1970s. If all variables are endogenous, then there is a serious identification problem with cross-sectional data that cannot be overcome no matter how much data are collected. For example, in the bivariate case where "causal thinking" may influence "behavioralism" as well as "behavioralism" influencing "causal thinking," the researcher only observes a single correlation which cannot produce the two distinctive coefficients representing the impact of "behavioralism" on "causal thinking" and the impact of "causal thinking" on "behavioralism."

The technical solution to this problem is the use of "instrumental variables" known to be exogenous and known to be correlated with the included endogenous variables, but the search for instruments proved elusive in many situations. Jackson (Chapter 17) summarizes the current situation with respect to "endogeneity and structural equation estimation" through his analysis of a simultaneous model of electoral support and congressional voting records. Jackson's chapter covers a fundamental problem with grace and lucidity, and he is especially strong in discussing "Instrumental Variables in Practice" and tests for endogeneity. Jackson's observations on these matters are especially appropriate because he was a member of the group that contributed to the 1973 Goldberger and Duncan volume on *Structural Equation Models in the Social Sciences*

date: 02 January 2017

which set the stage for several decades of work using these methods to explore causal relationships.

The most impressive accomplishment of this effort was the synthesis of factor analysis and causal modeling to produce what became known as LISREL, covariance structure, path analysis, or structural equation models. Bollen, Rabe-Hesketh, and Skrondal (Chapter 18) summarize the results of these efforts which typically used factor analysis types of models to develop measures of latent concepts which were then combined with causal models of the underlying latent concepts. These techniques have been important on two levels. At one level they simply provide a way to estimate more complicated statistical models that take into account both causal and measurement issues. At another level, partly through the vivid process of preparing "path diagrams," they provide a metaphor for understanding the relationships between concepts and their measurements, latent variables and causation, and the process of going from theory to empirical estimation. Unfortunately, the models have also sometimes led to baroque modeling adventures and a reliance on linearity and additivity that at once complicates and simplifies things too much. Perhaps the biggest problem is the reliance upon "identification" conditions that often require heroic assumptions about instruments.

One way out of the instrumental variables problem is to use time-series data. At the very least, time series give us a chance to see whether a putative cause "jumps" before (p. 20) a supposed effect. We can also consider values of variables that occur earlier in time to be "predetermined"—not quite exogenous but not endogenous either. Pevehouse and Brozek (Chapter 19) describe time-series methods such as simple time-series regressions, ARIMA models, vector autoregression (VAR) models, and unit root and error correction models (ECM). There are two tricky problems in this literature. One is the complex but tractable difficulty of autocorrelation, which typically means that time series have less information in them per observation than cross-sectional data and which suggest that some variables have been omitted from the specification (Beck and Katz 1996; Beck 2003). The second is the more pernicious problem of unit roots and commonly trending (co-integrated) data which can lead to nonsense correlations. In effect, in time-series data, time is almost always an "omitted" variable that can lead to spurious relationships which cannot be easily (or sensibly) disentangled by simply adding time to the regression. And thus, the special adaptation of methods designed for these data.

For our running example, we estimate a time-series autoregressive model for eighteen five-year periods from 1910 to 1999. The model regresses the proportion of articles mentioning "causal thinking" on the lagged proportions mentioning the words "behavioral or behavior," "regression," or "correlation." Table 1.4 shows that mentions of "correlation" do not seem to matter (the coefficient is negative and the standard error is bigger than the coefficient), but mentions of "regression" or "behavioralism" are

date: 02 January 2017

substantively large and statistically significant. (Also note that the autoregressive parameter is insignificant.) These results provide further evidence that it might have been the combination of behavioralism and regression that led to an increase in causal thinking in political science.

A time series often throws away lots of cross-sectional data that might be useful in making inferences. Time-series cross-sectional (TSCS) methods try to remedy this problem by using both sorts of information together. Beck (Chapter 20) summarizes this literature nicely. Not surprisingly, TSCS methods encounter all the (p. 21) problems that beset both cross-sectional and time-series data. Beck starts by considering the time-series properties including issues of nonstationarity. He then moves to cross-sectional issues including heteroskedasticity and spatial autocorrelation. He pays special attention to the ways that TSCS methods deal with heterogeneous units through fixed effects and random coefficient models. He ends with a discussion of binary variables and their relationship to event history models which are discussed in more detail in Golub (Chapter 23).

Table 1.4. Mentions of "causal thinking" for five-year periods regressed on mentions of "behavioral or behavior," "regression," and "correlation" for five-year periods for 1910–1999

Independent variables lagged	Regression coefficients (standard errors)
Behavior	.283 (.065)**
Regression	.372 (.098)*
Correlation	−.159 (.174)
AR (1)	.276 (.342)
Constant	−.002 (.005)
N	17 (one dropped for lags)

Significant: .05 ((*)) ,

.01 ((**)) ,

.001 ((***)) .

date: 02 January 2017

Martin (Chapter 21) surveys modern Bayesian methods of estimating statistical models. Before the 1990s, many researchers could write down a plausible model and the likelihood function for what they were studying, but the model presented insuperable estimation problems. Bayesian estimation was often even more daunting because it required not only the evaluation of likelihoods, but the evaluation of posterior distributions that combined likelihoods and prior distributions. In the 1990s, the combination of Bayesian statistics, Markov Chain Monte Carlo (MCMC) methods, and powerful computers provided a technology for overcoming these problems. These methods make it possible to simulate even very complex distributions and to obtain estimates of previously intractable models.

Using the methods in this chapter, we could certainly estimate a complex time- series cross-sectional model with latent variable indicators for the rise of causal thinking in the social sciences. We might, for example, gather yearly data from 1940 onwards on our various indicators for six different political science journals that have existed since then.[12] We could collect yearly indicators for each latent variable that represents a concept (e.g. "causal" or "causality" for "causal thinking" and "behavior" or "behavioral" for "behavioralism"). We could postulate some time-series cross- sectional model for the data which includes fixed effects for each journal and lagged effects of the explanatory variables. We might want to constrain the coefficients on the explanatory variables to be similar across journals or allow them to vary in some way. But we will leave this task to others.

6 Quantitative Tools for Causal and Descriptive Inference: Special Topics

Often our research requires that we use more specially defined methods to answer our research questions. In our running example, we have so far ignored the fact that our dependent variable is sometimes a dichotomous variable (as in Table 1.3 above), but there are good reasons to believe that we should take this into account. Discrete (p. 22) choice modeling (Chapter 22) by Glasgow and Alvarez presents methods for dealing with dichotomous variables and with ordered and unordered choices. These methods are probably especially important for our example because each journal article that we code represents a set of choices by the authors which should be explicitly modeled. Alvarez and Glasgow take readers to the forefront of this methodological research area by discussing how to incorporate heterogeneity into these models.

date: 02 January 2017

Golub's discussion of survival analysis (Chapter 23) presents another way to incorporate temporal information into our analysis in ways that provide advantages similar to those from using time series. In our running example, we could consider when various journals began to publish significant numbers of articles mentioning "causality" or "causal" to see how these events are related to the characteristics of the journals (perhaps their editorial boards or editors) and to characteristics of papers (such as the use of regression or behavioral language). As well as being a useful way to model the onset of events, survival analysis, also known as event history analysis, reveals the close ties and interaction that can occur between quantitative and qualitative research. For example, Elliott (2005) brings together narrative and event history analysis in her work on methodology.

A statistical problem that has commanded the attention of scholars for over a hundred years is addressed by Cho and Manski (Chapter 24). Scholars face this problem of "cross-level inference" whenever they are interested in the behavior of individuals but the data are aggregated at the precinct or census tract level. Cho and Manskid's chapter lays out the main methodological approaches to this problem; they do so by first building up intuitions about the problem. The chapter wraps up by placing the ecological inference problem within the context of the literature on partial identification and by describing recent work generalizing the use of logical bounds to produce solutions that are "regions" instead of point estimates for parameters.

The chapters on spatial analysis (Chapter 25) by Franzese and Hays and hierarchical modeling (Chapter 26) by Jones point to ways we can better capture the spatial and logical structure of data. In our running example, the smallest data unit was the use of words such as "causality" within the article, but these articles were then nested within journals and within years (and even in some of our analysis, within different disciplines). A complete understanding of the development of "causal thinking" within the sciences would certainly require capturing the separate effects of years, journals, and disciplines. It would also require understanding the interdependencies across years, journals, and disciplines.

Franzese and Hayes consider the role of "spatial interdependence" between units of analysis by employing a symmetric weighting matrix for the units of observation whose elements reflect the relative connectivity between unit i and unit j. By including this matrix in estimation in much the same way that we include lagged values of the dependent variable in time series, we can discover the impact of different forms of interdependence. In our example, if we had separate time series for journals, we could consider the impact of the "closeness" of editorial boards within disciplines based upon overlapping membership or overlapping places of training. These interdependencies could be represented by a "spatial" weighting matrix whose entries (p. 23) represent the degree of connection between the journals. The inclusion of this matrix in analyses poses

date: 02 January 2017

a number of difficult estimation problems, but Franzese and Hayes provide an excellent overview of the problems and their solutions.

Jones considers multilevel models in which units are nested within one another. The classic use of multilevel models is in educational research, where students are in classrooms which are in schools which are in school districts that are in states. Data can be collected at each level: test scores for the students, educational attainment and training for the teachers, student composition for the schools, taxing and spending for the school districts, and so forth. Multilevel methods provide a way of combining these data to determine their separate impacts on outcome variables.

At the moment, spatial and multilevel information cannot be easily incorporated in all types of statistical models. But these two chapters suggest that progress is being made, and that further innovations are on the way.

7 Qualitative Tools for Causal Inference

Throughout this chapter, we have been using our qualitative knowledge of American political science to make decisions regarding our quantitative analysis. We have used this knowledge to choose the time period of our analysis, to choose specific journals for analysis, to name our concepts and to select the words by which we have measured them by searching in JSTOR, to think about our model specifications, and to interpret our results. Now we use qualitative thinking more directly to further dissect our research problem.

Levy (Chapter 27) suggests that counterfactuals can be used along with case studies to make inferences, although strong theories are needed to do this. He argues that game theory is one (but not the only) approach that provides this kind of theory because a game explicitly models all of the actors' options including those possibilities that are not chosen. Game theory assumes that rational actors will choose an equilibrium path through the extensive form of the game, and all other routes are considered "off the equilibrium path"—counterfactual roads not taken. Levy argues that any counterfactual argument requires a detailed and explicit description of the alternative antecedent (i.e. the cause which did not occur in the counterfactual world) which is plausible and involves a minimal rewrite of history, and he suggests that one of the strengths of game theory is its explicitness about alternatives. Levy also argues that any counterfactual argument requires some evidence that the alternative antecedent would have actually led to a world in which the outcome is different from what we observe with the actual antecedent.

date: 02 January 2017

(p. 24) Short of developing game theory models to understand the history of political science, Levy tells us that we must at least try to specify some counterfactuals clearly to see what they might entail. One of our explanations for the rise of "causal thinking" is the invention of regression. Hence, one counterfactual is that regression analysis is not invented and therefore not brought into political science. Would there be less emphasis on causality in this case? It seems likely. As noted earlier, regression analysis, much more than correlation analysis, provides a seductive technology for exploring causality. Its asymmetry with a dependent variable that depends on a number of independent variables lends itself to discussions of causes (independent variables) and effects (dependent variables), whereas correlation (even partial correlation) analysis is essentially symmetric. Indeed, path analysis uses diagrams which look just like causal arrows between variables. Econometricians and statisticians provide theorems which show that if the regression model satisfies certain conditions, then the regression coefficients will be an unbiased estimate of the impact of the independent variables on the dependent variables. Regression analysis also provides the capacity to predict that if there is a one-unit change in some independent variable, then there will be a change in the dependent variable equal to the value of the independent variable's regression coefficient. In short, regression analysis delivers a great deal whereas correlation analysis delivers much less.

Yet, it is hard to believe that regression analysis would have fared so well unless the discipline valued the discussion of causal effects—and this valuation depended on the rise of behavioralism in political science to begin with. It seems likely that be- havioralism and regression analysis complemented one another. In fact, if we engage in a counterfactual thought experiment in which behavioralism does not arise, we speculate that regression alone would not have led to an emphasis on causal thinking. After reflection, it seems most likely that behavioralism produced fertile ground for thinking about causality. Regression analysis then took advantage of this fertile soil to push forward a "causal modeling" research agenda.[13]

It would be useful to have some additional corroboration of this story. With so many journal articles to hand in JSTOR, it seems foolhardy not to read some of them, but how do we choose cases? We cannot read all 78,046 articles from 1910 to 1999. Gerring (Chapter 28) provides some guidance by cataloging nine different techniques for case selection: typical, diverse, extreme, deviant, influential, crucial, pathway, most similar, and most different. Our judgment is that we should look for influential, crucial, or pathway cases. Influential cases are those with an influential configuration of the independent variables. Gerring suggests that if the researcher (p. 25) is starting from a quantitative database, then methods for finding influential outliers can be used. Crucial cases are those most or least likely to exhibit a given outcome. Pathway cases help to illuminate the mechanisms that connect causes and effects.

date: 02 January 2017

To investigate the role of behavioralism, we chose a set of four cases (sorted by JSTOR's relevance algorithm) that had "behavioralism" or "behavioral" in their titles or abstracts and that were written between 1950 and 1969. We chose them on the grounds that they might be pathway cases for behavioralism. The first article, by John P. East (1968), is a criticism of behavioralism, but in its criticism it notes that the behavioralist's "plea for empirical or causal theory over value theory is well known" (601) and that behavioralism "employs primarily empirical, quantitative, mathematical, and statistical methods" (597). The second article by Norman Luttbeg and Melvin Kahn (1968) reports on a survey of Ph.D. training in political science. The data are cross-tabulated by "behavioral" versus "traditional" departments with the former being much more likely to offer "behavioral" courses on "Use and Limits of Scientific Method" (60 percent to 20 percent), "Empirically Oriented Political Theory (60 percent to 24 percent), or "Empirical Research Methods" (84 percent to 48 percent) and being much more likely to require "Statistical Competence" (43 percent to 4 percent). The third article ("The Role for Behavioral Science in a University Medical Center") is irrelevant to our topic, but the fourth is "A Network of Data Archives for the Behavioral Sciences" by Philip Converse (1964). Converse mentions regression analysis in passing, but the main line of his argument is that with the growing abundance of survey and other forms of data and with the increasing power of computers, it makes sense to have a centralized data repository. The effort described in this article led to the ICPSR whose fortunes are reviewed in a later chapter in this handbook. After reading these four cases, it seems even more likely to us that behavioralism came first, and regression later. More reading might be useful in other areas such as "causal modeling" or "regression analysis" during the 1970s.

Rathbun (Chapter 29) offers still another method for understanding phenomena. He recommends intensive, in-depth interviews which can help to establish motivations and preferences, even though they must deal with the perils of "strategic reconstruction." Certainly it seems likely that interviews with those who lived through the crucial period of the 1950s to the 1970s would shed light on the rise of causal thinking in political science. Lacking the time to undertake these interviews, two of us who are old enough to remember at least part of this period offer our own perspectives. We both remember the force with which statistical regression methods pervaded the discipline in the 1970s. There was a palpable sense that statistical methods could uncover important causal truths and that they provided political scientists with real power to understand phenomena. One of us remembers thinking that causal modeling could surely unlock causal mechanisms and explain political phenomena.

Andrew Bennett (Chapter 30) offers an overview of process tracing, understood as an analytic procedure through which scholars make fine-grained observations to test ideas about causal mechanisms and causal sequences. He argues that the logic of (p. 26)

date: 02 January 2017

process tracing has important features in common with Bayesian analysis: It requires clear prior expectations linked to the theory under investigation, examines highly detailed evidence relevant to those expectations, and then considers appropriate revisions to the theory in light of observed evidence. With process tracing, the movement from theoretical expectations to evidence takes diverse forms, and Bennett reviews these alternatives and illustrates them with numerous examples.

Benoît Rihoux (Chapter 31) analyzes the tradition of case-oriented configurational research, focusing specifically on qualitative comparative analysis (QCA) as a tool for causal inference. This methodology employs both conventional set theory and fuzzy-set analysis, thereby seeking to capture in a systematic framework the more intuitive procedures followed by many scholars as they seek to "make sense of their cases." Rihoux explores the contrasts between QCA procedures and correlation-based methods, reviews the diverse forms of QCA, and among these diverse forms presents a valuable discussion of what he sees as the "best practices."

Much of what we have been doing in our running example in this chapter is to try to fathom the course of history—albeit a rather small political science piece of it. Comparative historical analysis provides an obvious approach to understanding complicated, drawn-out events. Mahoney and Terrie (Chapter 32) suggest that comparative historical analysis is complementary to statistical analysis because it deals with "causes of effects" rather than "effects of causes." Whereas statistical analysis starts from some treatment or putative cause and asks whether it has an effect, comparative historical analysis tends to start with a revolution, a war, or a discipline concerned with causal analysis, and asks what caused these outcomes, just as a doctor asks what caused someone's illness. In some cases, these are singular events which pose especially difficult problems—for doctors, patients, and political science researchers.

After providing a diagnosis of the distinctive features of historical research, Ma- honey and Terrie go on to provide some ideas about how we can tackle the problems posed by engaging in comparative historical inquiry. In our case, it seems likely that some comparative histories of American and European political science might yield some insights about the role of behavioralism and regression analysis. Another comparative approach would be to compare articles in journals with different kinds of editorial boards. Figure 1.3 suggests that there are substantial differences in the growth of mentions of "causal thinking" in the *American Political Science Review (APSR)*, *Journal of Politics (JOP)*, and *Review of Politics (ROP)* between 1940 and 1999. It would be useful to compare the histories of these journals.

date: 02 January 2017
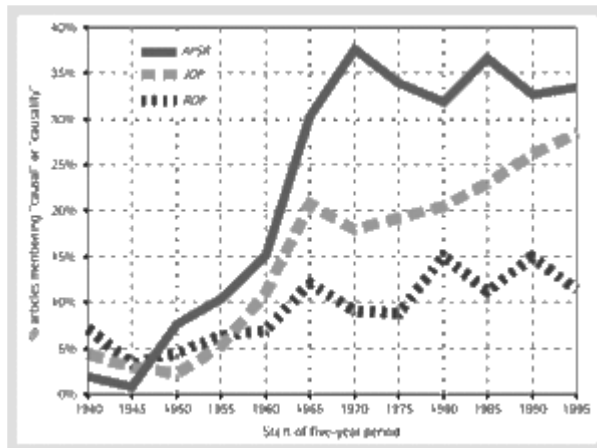
Click to view larger

Fig. 1.3. Growth of "causal thinking" in three journals 1940–1999

Fearon and Laitin (Chapter 33) discuss how qualitative and quantitative tools can be used jointly to strengthen causal inference. Large-n correlational analysis offers a valuable point of entry for examining empirical relationships, but if it is not used in conjunction with fully specified statistical models and insight into mechanisms, it makes only a weak contribution to causal inference. While case studies do not play a key role in ascertaining whether these overall empirical relations exist, they are valuable for establishing if the empirical relationships can be interpreted causally. Fearon and Laitin argue that this use of case studies will be far more valuable if the cases (p. 27) are chosen randomly. In our running example, this suggests that we should choose a number of articles in JSTOR at random and read them carefully. We might even stratify our sample so that we get more coverage for some kinds of articles than others.

8 Organizations, Institutions, and Movements in the Field of Methodology

If nothing else, the preceding pages should convince most people that organizations, institutions, and movements matter in political science. They certainly mattered for the behavioralists, and they have mattered for political methodologists. The final chapters review some of these movements—several of which involved the present authors at first hand.[14]

(p. 28) A clear trajectory in our discipline is that more and more attention is being devoted to methodology writ large. There is ample evidence for this assertion. The two methodology sections of the American Political Science Association are two of the largest of thirty-eight sections. There is an increasing demand for training in methodology. The discipline has expanded its ability to train its own graduate students (instead of sending them to economics or some other discipline), and there is an increasing capacity to better

train our undergraduates in methodology as well. Training is now available at the venerable Inter-University Consortium for Political and Social Research (ICPSR) Summer Training Program in methods, the Empirical Implications of Theoretical Models (EITM) summer programs that link formal models and empirical testing, and the winter Consortium on Qualitative Research Methods (CQRM) training program on qualitative methods. Methodology is taught more and more by political scientists to political scientists. Political methodology is also finding more and more connections with theory. Beck (2000) draws the contrast between statisticians and political methodologists in that "statisticians work hard to get the data to speak, whereas political scientists are more interested in testing theory." The focus on theory draws both quantitative and qualitative political scientists to the substance of politics, and it helps unite political methodologists to the political science community.

The rapid development of institutions for the study of qualitative methods in the past decade is discussed by Collier and Elman (Chapter 34). The discipline's welcoming response to these institutions reflects the pent-up need for them and the pluralistic culture of political science which facilitated the development of both the CQRM and the American Political Science Association's organized section on Qualitative Methods, recently renamed the Qualitative and Multi-Method Research Section.

Franklin (Chapter 35) traces the history of the quantitative methodology institutions, ICPSR, and the American Political Science Association's Political Methodology Section. ICPSR has the longest history, having been established in the 1960s in response to the needs of a newly quantitative field that lacked a tradition of training in statistical techniques. It was not until 1984 that the Political Methodology Section was formed to respond to the intellectual concerns driving the field.

Lewis-Beck (Chapter 36) discusses the forty-year history of publications in quantitative political methodology. He shows that the range and scope of outlets now available stands in dramatic contrast to what existed forty years ago.

Finally, Aldrich, Alt, and Lupia (Chapter 37) discuss the National Science Foundation's initiative to close the gap between theory and methods. The original goal of the Empirical Implications of Theoretical Models (EITM) initiative was to create a new generation of scholars who knew enough formal theory and enough about methods to build theories that could be tested, and methods that could test theories. Aldrich, Alt, and Lupia talk about the EITM as currently understood as a way of thinking (p. 29) about causal inference in service to causal reasoning. The empirical tool kit is seen as encompassing statistical approaches, experiments, and qualitative methods.

date: 02 January 2017

As Franklin rightly points out, academic institutions develop and are sustained because there are intellectual and professional needs that they serve. And these institutions matter. We know this as political scientists and see it in the development of our methodology field. Based on the vibrancy of our institutions, the future of political methodology looks bright indeed.

9 What Have We Learned?

The field of political methodology has changed dramatically in the past thirty years. Not only have new methods and techniques been developed, but the Political Methodology Society and the Qualitative and Multi-Method Research Section of the American Political Science Association have engaged in ongoing research and training programs that have advanced both quantitative and qualitative methodology. The *Oxford Handbook of Political Methodology* is designed to reflect these developments. Like other handbooks, it provides overviews of specific methodologies, but it also emphasizes three things.

■ *Utility for understanding politics*—Techniques should be the servants of improved data collection, measurement, and conceptualization and of better understanding of meanings and enhanced identification of causal relationships. The handbook describes techniques with the aim of showing how they contribute to these tasks, and the emphasis is on developing good research designs. The need for strong research designs unites both quantitative and qualitative research and provides the basis upon which to carry out high-quality research. Solid research design "… ensures that the results have internal, external, and ecological validity" (Educational Psychology).

■ *Pluralism of approaches*—There are many different ways that these tasks can be undertaken in the social sciences through description and modeling, case-study and large-n designs, and quantitative and qualitative research.

■ *Cutting across boundaries*—Techniques can and should cut across boundaries and should be useful for many different kinds of researchers. For example, in this handbook, those describing large-n statistical techniques provide examples of how their methods inform, or may even be adopted by, those doing case studies or interpretive work. Similarly, authors explaining how to do comparative historical work or process tracing reach out to explain how it could inform those doing time-series studies.

Despite its length and heft, our volume does not encompass all of methodology. As we indicated earlier, there is a rich set of chapters contained in a companion volume, (p. 30)

date: 02 January 2017

the *Oxford Handbook of Contextual Political Analysis*. This volume discusses interpretive and constructivist methods, along with broader issues of situating alternative analytic tools in relation to an understanding of culture. The chapter by Mark Bevir insightfully addresses questions of meta-methodology, a topic explored more widely in the other volume in discussions of epistemology, ontology, logical positivism, and postmodernism. Another important focus in the other volume is narrative analysis, as both a descriptive and an explanatory tool. Finally, in the traditions of research represented in our volume, the issues of context that arise in achieving measurement validity and establishing causal homogeneity are of great importance. But, corresponding to its title—i.e. *contextual political analysis*—the companion volume offers considerably more discussion of context and contextualized comparison which can be seen as complementary to our volume.

We hope that our running example on American political science has shown that at least some research problems (and perhaps all of them) can benefit from the use of both quantitative and qualitative methods. We find that quantitative methods provide some important insights about the size and scope of phenomena and about the linkages among variables, but quantitative methods are often maddeningly opaque with respect to the exact causal mechanisms that link our variables. Qualitative methods fill in some of these dark corners, but they sometimes lead to worries about the possibility that we have simply stumbled across an idiosyncratic causal path. We find ourselves oscillating back and forth between the methods, trying to see if insights from one approach can be verified and explicated by the other. But both are certainly helpful.

With respect to our running example, we conclude, with some trepidation given the incompleteness of our analysis, that values and inventions both help explain the rise of "causal thinking" in political science. The behavioral movement furthered "scientific values" like causal thinking, and regression provided an invention that seemingly provided political scientists with estimates of causal effects with minimal fuss and bother. As this handbook shows, however, regression is not the philosopher's stone that can turn base observational studies into gold-standard experimental studies. And even experimental studies have their limits, so that we are forced to develop an armamentarium of methods, displayed in this handbook, for dragging causal effects out of nature and for explaining political phenomena.

References

ARROW, K. J. 1963. *Social Choice and Individual Values*, 2nd edn. New Haven, Cann.: Yale University Press.

BECK, N. 2000. Political methodology: a welcoming discipline. *Journal of the American Statistical Association*, 95: 651–4.

date: 02 January 2017

—— 2003. Time-series cross-section data: what have we learned in the past few years? *Annual Review of Political Science*, 4: 271–93.

(p. 31) —— and ᴋᴀᴛᴢ, J. N. 1996. Nuisance vs. substance: specifying and estimating time-series cross- section models. *Political Analysis*, 89: 634–47.

ᴄᴏɴᴠᴇʀsᴇ, P. 1964. A network of data archives for the behavioral sciences. *Public Opinion Quarterly*, 28: 273–86.

ᴇᴀsᴛ, J. P. 1968. Pragmatism and behavioralism. *Western Political Science Quarterly*, 21: 597–605.

ᴇʟʟɪᴏᴛᴛ, J. 2005. *Using Narrative in Social Research: Qualitative and Quantitative Approaches*. London: Sage.

ғɪsʜᴇʀ, R. A. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

ғʀᴇᴇᴅᴍᴀɴ, D. A. 2005. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.

ɢᴏʟᴅʙᴇʀɢᴇʀ, A. and ᴅᴜɴᴄᴀɴ, O. D. 1973. *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

ʜᴏʟʟᴀɴᴅ, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association*, 81: 945–60.

ᴋɪɴɢ, G. ᴋᴇᴏʜᴀɴᴇ, R. O. and ᴠᴇʀʙᴀ, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

ʟᴇᴀᴍᴇʀ, E. E. 1983. Let's take the con out of econometrics. *American Economic Review*, 73: 31–43.

ʟᴜᴛᴛʙᴇɢ, N. R. and ᴋᴀʜɴ, M. A. 1968. Ph.D. training in political science. *Midwest Journal of Political Science*, 12: 303–29.

ᴍᴄᴋᴇʟᴠᴇʏ, R. D. 1979. General conditions for global intransitivities in formal voting models. *Econometrica*, 47: 1085–112.

ɴᴇʏᴍᴀɴ, J. 1923. On the application of probability theory to agricultural experiments: essay on principles, **section 9**; trans. 1990. *Statistical Science*, 5: 465–80.

PEARSON, K. 1896. Mathematical contributions to the theory of evolution: III. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London*, 187: 253–318.

—— 1909. Determination of the coefficient of correlation. *Science*, 30: 23–5.

RUBIN, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66: 688–701.

—— 1978. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6: 34–58.

SAARI, D. G. 1999. Chaos, but in voting and apportionments? *Proceedings of the National Academy of Sciences of the United States of America*, 96: 10568–71.

SCHOFIELD, N. J. 1983. Generic instability of majority rule. *Review of Economic Studies*, 50: 695–705.

YULE, G. U. 1907. On the theory of correlation for any number of variables, treated by a new system of notation. *Proceedings of the Royal Society of London*, 79: 182–93. (p. 32)

Notes:

(1) If we just search for the words "cause" or "causes" alone in all political science articles, we find that the proportion of these words is 55 percent in 1995–9 which is not a very dramatic increase since 1910–19 when it was 50 percent. This suggests that the words "cause" or "causes" measure something different from "causality" and "causal." As we shall see, political methodology often grapples with questions like this about construct validity.

(2) We might also search for the term "least squares" but almost whenever it appears (88 percent of the time), the term "regression" also appears, so not much is gained by searching for it as well.

(3) Using a list of presidents of the American Political Science Association, we coded those people known to be "behavioralists" from 1950 to 1980—we coded sixteen of the 31 presidents in this way (Odegard, Herring, Lasswell, Schattschneider, Key, Truman, Almond, Dahl, Easton, Deutsch, Lane, Eulau, Leiserson, Ranney, Wahlke, and Miller). Using different time periods yields similar results. (For example, the 1950–80 period yields 35 percent for a general article and 78 percent for those by the famous behavioralists.)

date: 02 January 2017

(4) We constructed variables for each word with a zero value if the word was not present in an article and a one if it was mentioned at least once. Then we obtained the ten correlations between pairs of the five variables with articles as the unit of analysis.

(5) Each word appears in a different number of articles, but one or the other or both of the words "narrative" or "interpretive" appear in about 5.9 percent of the articles and the words "hypothesis" or "causal" or "causality" appear in almost one-third (31.3 percent). "Explanation" alone appears in 35.4 percent of the articles.

(6) In 1980–4, the words "narrative" or "interpretive" were mentioned only 4.1 percent of the time in political science journals; in the succeeding five-year periods, the words increased in use to 6.1 percent, 8.1 percent, and finally 10.1 percent for 1995–9.

(7) At least two other words might be relevant: "law" and "theory." The first gets at the notion of the need for "law-like" statements, but searching for it on JSTOR obviously leads to many false positives—mentions of public laws, the rule of law, the study of law, and the exercise of law. Similarly, "theory" gets at the notion of "theories" lying behind hypotheses, but the subfield of "political theory" uses theory in a much different sense.

(8) Thus if C is cause and E is effect, a necessary condition for causality is that $\text{Prob}(E|C) > \text{Prob}(E|\text{not } C)$. Of course, this also means that the expectation goes up $E(E|C) > E(E|\text{not } C)$.

(9) Not until the 1960s are there any articles that use the term "regression" and either "causal model" or "causal modeling." Then the number grows from 25 in the 1960s, to 124 in the 1970s, to 129 in the 1980s. It drops to 103 in the 1990s.

(10) SUTVA means that a subject's response depends only on that subject's assignment, not the assignment of other subjects. SUTVA will be violated if the number of units getting the treatment versus the control status affects the outcome (as in a general equilibrium situation where many people getting the treatment of more education affects the overall value of education more than when just a few people get education), or if there is more communication of treatment to controls depending on the way assignment is done.

(11) The observant reader will note that these authors make a causal claim about the power of an invention (in this case experimental methods) to further causal discourse.

(12) *American Political Science Review* (1906), *Annals of the American Academy of Political and Social Science* (1890), *Journal of Politics* (1939), *Political Science Quarterly* (1886), *Public Opinion Quarterly* (1937), and *Review of Politics* (1939).

date: 02 January 2017

(13) The time-series analysis provides some support for this idea. If we regress the proportion of articles mentioning behavioralism on its lagged value and the lagged values of the proportion of articles mentioning regression, correlation, and causality, only behavioralism lagged has a significant coefficient and causality and correlation have the wrong signs. Behavioralism, it seems, is only predicted by its lagged value. If we do the same analysis by regressing causality on its lagged value and the lagged values of regression, correlation, and behavioralism, we find that only behavioralism is significant and correlation has the wrong sign. If we eliminate correlation, then causality has the wrong sign. If we then eliminate it, we are left with significant coefficients for behavioralism and regression suggesting that mentions of causality come from both sources.

(14) Brady was a founding member and early president of the Political Methodology Society. He was a co-principal investigator (with PI Paul Sniderman and Phil Tetlock) of the Multi-Investigator Study which championed the use of experiments in surveys and which provided the base for the TESS program. And he was present at the meeting convened by Jim Granato at NSF which conceived of the EITM idea, and he is a co-PI of one of the two EITM summer programs. Janet Box-Steffensmeier was an early graduate student member of the Political Methodology Society and a recent President. David Collier was the founding President of the APSA qualitative methods section, and the Chair of CQRM's Academic Council.

Janet M. Box-Steffensmeier

Janet M. Box-Steffensmeier is Vernal Riffe Professor of Political Science and Sociology, Director of the Program in Statistics Methodology, Ohio State University.

Henry E. Brady

Henry E. Brady is Professor of Political Science and Public Policy at University of California, Berkeley. He received his Ph.D. in Economics and Political Science from MIT, and his areas of interest include Quantitative Methodology, American and Canadian Politics, and Political Behavior. He teaches undergraduate courses on political participation and party systems and graduate courses on advanced quantitative methodology.

David Collier

David Collier is Professor of Political Science at University of California, Berkeley and former President of the American Political Science Association. His fields are comparative politics, Latin American politics, and methodology. His latest book is Rethinking Social Inquiry: Diverse Tools, Shared Standards, of which he is co-editor and co-author with Henry E. Brady.

date: 02 January 2017

date: 02 January 2017