POLITICAL SCIENCE 522
RESEARCH DESIGN AND ANALYSIS IN QUANTITATIVE RESEARCH
Spring 2014
1:00-3:30T

Overview

Instructor: James H. Kuklinski and Jake Bowers
Offices: 404 DKH and 213 IGPA (Jim) and 432 DKH (Jake)
Phone Numbers: 333-9589 (DKH) and 244-7983 (IGPA) (Jim) and no phone number for Jake
Office Hours: 2:00-3:30 W or by Appointment (Jim and Jake)

Political Science 522 begins with the premise that good question, good theory, good research design, and good writing all go hand-in-hand. No design will overcome a muddled question, nor will a poorly planned or executed design suffice to answer a clearly stated question. In the presence of poor writing, not much else matters. Political Science 522 thus focuses on the question-theory-design-writing connections, probably more than many research design courses do.

The content and organization of this course have changed markedly over the years, largely because political scientists, especially during the past decade, have been seriously rethinking the foundations of empirical research. Until about the end of the 1950s, political scientists would commonly undertake an in-depth field study of a phenomenon in a particular country or region of a country. The emphasis was detailed understanding of a specific context. A particularly powerful critique of this work was its limited capacity to reach conclusions that apply across units. This critique, combined with a growing availability of quantitative data and statistical methods to analyze them, led to the emergence of the statistical analysis of relatively large data sets. For more than half a century, political scientists would rely heavily on regression analysis and related methods to make general causal claims. In retrospect, it is now clear that the regression approach emphasized the "general" more than the "causal." Currently, more and more scholars are questioning the utility of this relative emphasis: "What good is generalization if the estimated causal effect is wrong?" (Of course, one can easily reverse the question: "What good is showing cause and effect if it cannot be generalized beyond the specific cases?" The syllabus keeps both questions close-at-hand.)

Four recent disciplinary trends have begun to leave their marks. First, political scientists have become much more aware of the specification problems that inhere in statistical models designed to analyze observational data. An extreme although not wholly wrong view is that a researcher simply can never be sure that he or she has properly specified the statistical model. Second, influential work by

statisticians has led political scientists to think of causality in terms of counterfactuals—"what if X had not occurred," or, alternatively, "what if X had occurred"—rather than in terms of associations—"what is the magnitude of the association between X and Y". Third, there has been a rapidly growing interest in finding ways to identify heterogeneous treatment effects in the data, rather than to be satisfied with an average, consistent-across-all-units treatment effect. Heterogeneity can occur for various reasons, including that the treatment affects some units differently from others and that some units are exposed to (expose themselves to) the treatment while others are (do) not. Finally, political scientists have become increasingly dissatisfied with "black-box" regression estimates, and thus begun to try to identify the mechanisms by which X influences Y. This seemingly straightforward task has proven to be difficult, with some prominent methodologists suggesting that it might be impossible to identify mechanisms in quantitative research.

Arguably, a broader trend has motivated all four of the trends just mentioned. This is the growing recognition that unobserved (not measured or not-able-to-be-measured) factors can seriously bias estimates of relationships. Indeed, a good part of the syllabus is devoted to distinguishing methods that select on observables from those that select on un-observables. Some of the discipline's most creative scholars have been working in the second vein.

The academic community defines what are and are not adequate explanations of political phenomena. Although the discipline's diversity precludes a consensus on this matter, graduate students who gain a solid understanding of the lay-of-the-land, including the changes that have occurred and will likely occur, will be best positioned to publish influential work.

The Nature and Challenges of Research Design

Research design is not statistics, even though the two are intertwined. Research design emphasizes the formulation of studies that produce convincing results. When one writes a paper and sends it out for review, the referees assigned to evaluate it will look for weaknesses: Has the author adequately considered alternative explanations? Has the author taken censoring and mutual causation into account? Does the author understand the process by which his or her data were generated? How does self-selection affect the interpretation of the statistical findings? Although a researcher might use statistics to assist in addressing such questions, the questions themselves are rooted in research design, which always precedes statistical analysis. Indeed, no amount of sophisticated statistical analysis can mend a poorly designed study.

Understand that there is no set recipe for "doing" a good research design, which requires bringing *the right mind state* to bear. At a minimum, the researcher must

learn how to anticipate and address criticisms that others will make. This course will identify some of the criticisms that are likely to arise; each student must then figure out how to make his or her research as "untouchable" as possible. This takes practice, lots and lots of practice. It also requires, as suggested above, a broad understanding of changes in disciplinary expectations.

Unfortunately, it is not simply a matter of the researcher identifying all potential criticisms. Any research design entails making tradeoffs: external versus internal validity, generalization versus detail and specificity, right theory and weak data versus slightly right theory and strong data, to name just a few. Students who truly enjoy research will find satisfaction in balancing the various tradeoffs, frustrations notwithstanding. Those who feel more comfortable following canned recipes will find less satisfaction in such intellectual engagement.

Two comments on the readings: First, many of the assigned books and articles easily fit into more than one subsection of the syllabus because they illustrate more than a single lesson. Thus, some readings are listed more than once. Returning to a reading from a different perspective can be illuminating. Second, we will not read every single item listed in the syllabus, in fact, will likely not read selections from every subsection. Some pieces are "must" reading; in other cases, we will pick and choose on the basis of students' research interests.

Course Goals

The overarching goal of this course is to increase sensitivity to the following:

The critical importance of good writing

The importance of knowing current disciplinary standards and expectations

The importance and challenge of proper question formulation

The connection between question formulation and research design

The core challenges in research design and analysis

Course Requirements

We will discuss course requirements at our first class meeting.

Rules and Guidelines

Only three rules apply to this course: all assignments will be turned in on time; there will be no incompletes; and students who fail to prepare fully before class need not attend, and should consider an alternative course.

Reading Materials

Most journal articles will be available online; otherwise the article will be placed in the departmental office for students to copy. Although we will not read every work listed below, the load will be heavy.

You can purchase the required books at Follett's or the Illini bookstore, or, if you prefer, via the Internet.

E. Carmines and R. Zaller, *Validity and Reliability Assessment*
W. Shadish, T. Cook and D. Campbell, *Experimental and Quasi-Experimental Designs*
G. King, R. Keohane, and S. Verba, *Designing Social Inquiry*
S. Lieberson, *Making It Count*
A. Garfinkel, *Forms of Explanation*
J. Miller and S. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*
H. Brady and D. Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*
J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*

As noted above, we will not read every item listed below. We will pick and choose once we know students' research interests.

A Few Words of Wisdom!

You might not agree with the following, in which case you undoubtedly will be able to offer a strong rationale for your disagreement(s).

"It is better for a social scientist to look smart than to be right."

"Matters of research design should give you pause, not paralyze you."

"Good research design will always trump sophisticated statistical analysis."

"Social scientists' work would be much easier if humans did not have intentions."

"The likely causal complexity of the world far outstrips the language that social scientists have available to understand it."

"That experiments are widely described as the gold standard for showing cause and effect and yet have played a secondary role in social scientific research until recently is inexplicable."

"The claim that social scientists 'do the best they can' holds little sway if 'the best they can' falls short of what is necessary to understand their worlds."

"A good writer who is a poor scientist will almost always be more successful as a social scientist than a good scientist who is a poor writer."

Fundamentals

The topics included in this section accompany all research studies. Even when a researcher does not explicitly address some of the topics, he or she will make decisions about them. Graduate students should view the topics in this section as a checklist of items to be addressed early in a study.

The Role of Comparison in Questions Asked and Questions Answered

Scientific discovery entails comparison. When posing a research question, the social scientist must state explicitly what comparison(s) he or she intends to make. Asking the question so that the right comparisons are made proves far more challenging than scholars are wont to believe. Garfinkel demonstrates how contrast spaces can be used to explicate precisely the question that is being asked and the presumptions that underlay it. An old book by academic standards, *Forms of Explanation* remains the single best discussion of research questions. Angrist and Pischke present a more modern (and limited) definition of a good research question. Their conception of proper comparison derives from experimental logic. Whether there is a tradeoff between breadth of research question and capacity to show a causal effect warrants thought.

Social scientists also make comparisons when they answer questions with evidence. Both Gaines and Kuklinski and Sekhon and Titiunik offer examples of where the authors of highly influential experimental studies were not making the comparisons they thought they were making. That the authors of the critiqued studies are among the best students of American politics underlines how easy it is inadvertently to make a wrong comparison. Wand et al.'s study of voting returns in the 2000 presidential

election represents an effective use of a variety of comparisons to reach a conclusion. It nicely illustrates how using a variety of comparisons can comprise a compelling argument.

A. Garfinkel, *Forms of Explanation*, Introduction, chs. 1-4
J. Angrist and J-P. Pischke, *Mostly Harmless Econometrics,* ch. 1
J. Wand, K. Shotts, J. Sekhon, W. Mebane, M. Herron, and H. Brady, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," *American Political Science Review*, 2001, pp. 793-810
B. Gaines and J. Kuklinski, "Incorporating Self-Selection into the Random Assignment Experiment," *American Journal of Political Science*, 2011, pp. 724-736
J. Sekhon and R. Titiunik, "When Natural Experiments Are Neither Natural nor Experiments," *American Political Science Review*, 2012, pp. 35-57 (focus on pp. 35--37 and pp. 46--51)

Theory and Evidence

At least three distinct and not necessarily mutually exclusive views of the relationship between theory and evidence exist. For the hardcore empiricist, theory means hypotheses that often come directly from the data. For others, the purpose of evidence is to test independently formulated (axiomatic) theory. For yet others, (axiomatic) theory is essential to understand the data-generating process, and thus to interpret the data themselves.

Sutton and Staw state, in emphatic terms, what theory is not, while Lave and March offer an elementary introduction to theory (more precisely, modeling) and suggest that the researcher move back and forth between model and data, while at the same time increasing the model's generality. Smith, Signorino and Yilmaz, and Bueno de Mesquita all argue why the analysis of data in the absence of rigorous theory is risky. Smith, in particular, demonstrates the value of game theory to understand the process that generated the available data. Anyone who studies political phenomena that include strategic behavior among actors should keep the lesson in Smith close-at-hand. Svolik's and Carter's studies are convincing examples of how to connect axiomatic theory and quantitative data analysis. In contrast, Clarke and Primo, in their recent and controversial book, argue that empirical and theoretical research should proceed independently.

Pahre and Achen and Snidal demonstrate the natural and often overlooked connection between theory and case studies. Rogowski argues that a

single case can suffice as ample evidence in the presence of good theory. Does he make a convincing case?

Most of the lessons of this subsection apply more directly to data generated by the real world than to data generated by the researcher via experiments, where the researcher typically has more control. However, the lessons are sufficiently general to apply to both types of data.

R. Sutton and B. Staw, "What Theory is not," *Administrative Science Quarterly*, 1995, pp. 371-384
C. Lave and J. March, *An Introduction to Models in the Social Sciences*, pp. 1-84
A. Smith, "Testing Theories of Strategic Choice: The Example of Crisis Escalation," *American Journal of Political Science,* 1999, pp. 1254-1283
K. Signorino and K. Yilmaz, "Strategic Misspecification in Regression Models," *American Journal of Political Science,* 2003, pp. 551-566
E. Bueno de Mesquita, "The Quality of Terror," *American Journal of Political Science*, 2005, pp. 515-530
M. Svolik, "Power Sharing and Leadership Dynamics in Authoritarian Regimes," *American Journal of Political Science*, 2009, pp. 477-494
D. Carter, "The Strategy of Territorial Conflict," *American Journal of Political Science*, 2010, pp. 969-987
K. Clarke and D. Primo, *A Model Discipline*, all
R. Pahre, "Formal Theory and Case Study Methods in EU Studies," *European Union Politics*, 2004, pp. 113-146
C. Achen and D. Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics*, 1989, pp. 144-169
R. Rogowski, "How Inference Neglects Theoretical Anomaly," in H. Brady and D. Collier, eds., *Rethinking Social Inquiry*

Research Question and Levels and Units of Analysis

It is temptingly easy for a researcher to work at a particular level of analysis without conscious consideration of why he or she chose that level. In fact, it is perilously easy to conduct research at a level of analysis that is inappropriate to the research question. Methodological individualists might be especially guilty in this regard. The classic exchange between Seligson and Inglehart and Welzel illustrates that the "right" level of analysis might not be obvious, even upon reflection. Which of the two pieces makes the more convincing argument? Why? Similarly, does Baek choose an appropriate level of analysis? Why or why not?

Four of the readings address the ecological inference problem, as first articulated by Robinson. Prior to the publication of Robinson's classic, political scientists routinely conducted their analyses at the level of the state, county, or community.

Robinson brought that practice to an abrupt halt, and even today, social scientists shy away from the use of aggregated data. King proposed a solution, which Cho, and Cho and Gaines, showed was no solution at all. Yet, the opportunities for political scientists to undertake research would be considerably higher if they could use aggregated data. So, the question is: are there circumstances where the use of aggregated data might be justifiable?

The next six readings all address, in one way or another, the task of spanning levels of analysis. It is especially useful to compare and contrast Steenbergen and Jones, Maoz, and Miller and Page (and Jung and Lake). The three approaches differ markedly, and yet each is valid in its own right. The following two studies, both very recent, illustrate how new technologies have improved the capacity of researchers to conduct analyses at appropriate levels of analysis.

The last three readings, in combination, focus on *unit* of analysis more than *level* of analysis. The Mondak et al. *American Political Science Review* article represents the best work on personality and politics. What is the unit of analysis with respect to question formulation? What is the unit of analysis with respect to data analysis? What are the equivalent units of analysis in the Pickles et al. study? How would the analysis (and perhaps question) change if Mondak et al. used latent class analysis rather than regression analysis? Does the choice of method/unit of analysis matter?

M. Seligson, "The Renaissance of Political Culture or the Renaissance of the Ecological Fallacy?" *Comparative Politics*, 2002, pp. 273-292
R. Inglehart and C. Welzel, "Political Culture and Democracy: Analyzing Cross-level Linkages," *Comparative Politics*, 2003, pp. 61-79
M. Baek, "A Comparative Analysis of Political Communication Systems and Voter Turnout," *American Journal of Political Science*, 2009, pp. 376-393
W.S. Robinson, "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, 1950, pp. 351-357
C. Achen and P. Shively, *Cross-Level Inference*, ch. 1
G. King, *A Solution to the Ecological Inference Problem*, ch. 1
W. Cho and B. Gaines, "The Limits of Ecological Inference: The Case of Split-Ticket Voting," *American Journal of Political Science*, 2004, pp. 152-171
D. Diermeier and K. Krehbiel, "Institutionalism as a Methodology," *Journal of Theoretical Politics*, 2003, pp. 123-144
J. Fowler, "Connecting the Congress: A Study of Co-sponsorship Networks," *Political Analysis*, 2006, pp. 456-487
Z. Maoz, "The Effects of Strategic and Economic Interdependence on International Conflict across Levels of Analysis, *American Journal of Political Science*, 2009, pp. 223-240
M. Steenbergen and B. Jones, "Modeling Multilevel Data Structures," *American Journal of Political Science*, 2002, pp. 218-237

M. Macy and R. Willer, "From Factors to Actors: Computational Sociology and Agent-Based Modeling," *Annual Review of Sociology*, 2002, pp. 143-166
J. Miller and S. Page, *Complex Adaptive Systems*, chs. 1-10
D. Jung and D. Lake, "Markets, Hierarchies, and Networks: An Agent-Based Organizational Ecology," *American Journal of Political Science*, 2011, pp. 971-989
M. Nepal, A. Bohara, and K. Gawande, "More Inequality, More Killings: The Maoist Insurgency in Nepal, *American Journal of Political Science*, 2011, pp. 885-905
L. Cederman, N. Weidman, and K. Gleditsch, "Horizontal Inequalities and Ethno-Nationalist Civil War: A Global Comparison," *American Political Science Review*, 2011, pp. 478-495
J. Mondak et al., "Personality and Civic Engagement: An Integrative Framework for the Study of Trait Effects on Political Behavior," *American Political Science Review*, 2010, pp. 85-110
A. Linzer and J. Lewis, "**poLCA**: Polytomous Variable Latent Class Analysis," Version 1.2, Emory University
A. Pickles, P. Bolton, H. MacDonald, A. Bailey, A. Le Couteur, C. Sim, and M. Rutter, "Latent-Class Analysis of Recurrence Risks for Complex Phenotypes with Selection and Measurement Error: A Twin and Family History Study of Autism," *American Journal of Genetics*, 1995, pp. 717-726

Concepts Defined, Concepts Measured

Researchers communicate with each other primarily through concepts. Indeed, concepts define an individual researcher's identity far more than theory or data do. When researchers choose to use some concepts and not others, they identify their relevant scholarly communities. As academic research continues to become more and more specialized, the more will scholars need to choose their concepts carefully and selectively?

Unfortunately, social scientists emphasize data analysis and variables far more than the proper formulation and measurement of concepts. If researchers do not carefully formulate concepts and then validly and reliably measure them, sophisticated analysis cannot save the day. Nor will useful communication across scholars be possible.

The following readings are divided into four categories: general treatments of concept definition, general treatments of measurement, specific examples, and missing data. Conceptualization and measurement can easily comprise a semester-long course, so, unfortunately, we will only scratch the surface here.

Kaplan's two chapters are the classic statement on concept formation. David Collier, Gary Goertz, and Charles Ragin have given more thought to concepts

than most other contemporary social scientist. Collier and his students approach concept formation from a comparative perspective, and ask how concepts can be defined to travel across countries and other units that differ markedly. Although a student of international relations, Goertz discusses concepts broadly. Ragin now believes in "fuzzy" concepts, an interesting idea that political scientists have been slow to adopt.

The next set of readings discusses measurement. Shadish et al. and Jackman present overviews of the challenges in measurement. Bartholomew questions whether constructs such as attitudes can be measured at all. Carmines and Zaller's monograph presents the basics of validity and reliability. Campbell and Fiske's discussion of the multitrait-multimethod technique offers one path to valid measurement. It effectively captures the logic of validity and reliability. Lieberson's chapter and Jacoby's article have implications for measurement. Derive those implications. Imai and Yamamoto show that differential measurement error can generate biased causal estimates.

Immediately following are 16 readings, selected from the discipline's three empirically-oriented fields, where the authors try explicitly to define concepts clearly and precisely and then measure them. These articles share a message: proper concept formation and measurement are essential and difficult. Poor measurement, in particular, tends to attenuate the strength of the relationship between an independent and dependent variable. The exchange between Sniderman and Tetlock on the one hand and Sears and Kinder illustrates how problems can arise when social scientists poorly define their concepts. The Gould chapters demonstrate, dramatically, that the hard sciences also face measurement problems, with perhaps more severe consequences.

Finally, one of the biggest pains in the butt is the lack of complete data. Survey data are the most notorious in this regard, but they are not alone. Most of the time, researchers ignore missing data. Rarely do they report the number of cases they dropped because the units had missing values on one or more variables. Thus the reader cannot assess whether the missing data are systematically biasing the reported results. In other words, missing data can befuddle efforts to measure concepts correctly. The imputation of missing values has, rightly, become a hot topic and warrants discussion. Allison, Howell, and King et al. offer their favorite algorithms for imputing missing data. Many now exist.

A. Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science*, chs. 1-2
D. Collier and J. Mahoney, "Conceptual 'Stretching' Revisited: Alternative Views of Categories in Comparative Analysis," *American Political Science Review*, 1993, pp. 845-855

R. Adcock and D. Collier, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review*, 2001, pp. 529-546
G. Goertz, *Social Science Concepts: A User's Guide*, all
G. Goertz, "Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*
C. Ragin, *Fuzzy-Set Social Science*, ch. 6
W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, pp. 64-82, 341-373
S. Jackman, "Measurement," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*
D. Bartholomew, "Scaling Unobservable Constructs in Social Science," *Applied Statistics*, 1998, pp. 1-13
E. Carmines and R. Zaller, *Reliability and Validity Assessment*, all (including appendix)
D. Campbell and D. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 1959, pp. 81-105
W. Jacoby, "Levels of Measurement and Political Research: An Optimist's View," *American Journal of Political Science*, 1999, pp. 271-301
S. Lieberson, *Making It Count*, ch. 5
K. Imai and T. Yamamoto, "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis," *American Journal of Political Science*, 2010, pp. 543-560

D. Laitin, "What is a Language Community?" *American Journal of Political Science*, 2000, pp. 142-155
Z. Elkins, "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations," *American Journal of Political Science*, 2000, pp. 293-300
J. Cheibub, J. Gandhi, and J. Vreeland, "Democracy and Dictatorship Revisited," Working paper, 2009
G. Munck and J. Verkuilen, "Conceptualizing and Measuring Democracy: Evaluating Alternative Indexes," *Comparative Political Studies*, 2002, pp. 5-34
R. Merritt and D. Zinnes, "Validity of Power Indices," *International Interactions*, 1988, pp. 141-153
G. Goertz and P. Diehl, "Enduring Rivalries: Theoretical Constructs and Empirical Patterns," *International Studies Quarterly*, 1993, pp. 147-171
J. Mondak, "Developing Valid Knowledge Scales," *American Journal of Political Science*, 2001, pp. 224-238
M. Laver and J. Garry, "Estimating Policy Positions from Political Texts," *American Journal of Political Science*, 2000, pp. 619-634

S. Ansolabehere, J. Rodden, and J. Snyder, Jr., "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting, *American Political Science Review*, 2008, pp. 215-232
D. Kinder and D. Sears, "Prejudice and Politics: Symbolic Racism versus Racial Threats to the Good Life," *Journal of Personality and Social Psychology*, 1981, pp. 414-431
P. Sniderman and P. Tetlock, "Symbolic Racism: Problems of Political Motive Attribution," *Journal of Social Issues*, 1986, pp. 129-150
D. Kinder, ""The Continuing American Dilemma: White Resistance to Racial Change 40 Years after Myrdal," *Journal of Social Issues*, 1986, pp. 151-172
P. Sniderman and P. Tetlock, "Reflections on American Racism," *Journal of Social Issues*, 1986, pp. 173-178
G. King, C. Murray, J. Salomon, and A. Tandon, "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," *American Political Science Review*, 2003, pp. 567-584
L. Bartels, "Messages Received: The Political Impact of Media Exposure," *American Political Science Review*, 1993, pp. 267-285
D. Green, S. Goldman, and P. Salovey, ""Measurement Error Masks Bipolarity in Affect Ratings," *Journal of Personality and Social Psychology*, 1993, pp. 1029-1041
H. Asher, "Some Consequences of Measurement Error in Survey Data," *American Journal of Political Science*, 1974, pp. 469-485
J. Wand, "Credible Comparisons Using Interpersonally Incomparable Data: Nonparametric Scales with Anchoring Vignettes," *American Journal of Political Science*, 2012, pp. 1-14
S. Dilliplane, S. Goldman, and D. Mutz, "Televised Exposure to Politics: New Measures for a Fragmented Media Environment," *American Journal of Political Science*, 2012, pp. 1-13
S. Gould, *The Mismeasure of Man*, chs. 3-4

P. Allison, *Missing Data*, all
D. Howell, "Treatment of Missing Data," unpublished paper
P. Allison, "Multiple Imputation for Missing Data: A Cautionary Tale," Unpublished paper
G. King et al., "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review*, 2001, pp. 49-69

Causal Analysis

Most social scientists adopt the assumption that "the most interesting research in social science is about questions of cause and effect" (Angrist and Pischke, *Mostly Harmless Econometrics*, p. 3). As graduate students quickly learn, social scientists' capacities to show cause and effect fall short of the task's importance.

Conceptions of Explanation and Causality

Social scientists seek to understand why and how the real-world phenomena that interest them work the way they do. They pose and try to answer "why" and "how" questions. Somewhere along the line, "explanation" became equated with "causal explanation." Note that "causal" serves as an adjective in the preceding term, which implies it is one kind of explanation, but not the only kind. This raises the question, what is explanation? To answer the question, we take a peek at what philosophers of science tell us. We read pieces from Hempel and Oppenheim, who summarize the logical positivist perspective, Salmon, who ponders the association perspective, and Scriven, who proposes an everyday, common-sense definition of explanation.

Scholars have offered numerous conceptions of causality. Heckman's 2005 article is far-ranging and worth reading. Holland, Fearon, and King, Keohane, and Verba adopt a similar and widely-held conception of causality. The other authors offer conceptions of causality that vary considerably. The offerings cover a wide range, from INUS (Mackie) to Boolean algebra (Ragin) to necessary and sufficient conditions (Braumoeller and Goertz) to path dependence (Bennett and Elman). Do some conceptions seem more compelling than others? On what basis might one choose one conception or test rather than others? How might selection of a conception shape question formulation and vice-versa?

The exchange between Ashworth et al. and Pape underlines a fundamental fact of causal analysis that we highlighted earlier: comparison is always a big part. Thus, a researcher needs to know what is and is not being compared; and, if in fact, whether something worthwhile is being compared.

In two recent and potentially influential articles, Yamamoto distinguishes between causal attribution and causal effects and Blackwell proposes a way to think about causal inference in dynamic rather than one-shot terms.

C. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science*, 1948, introduction, parts I and III
W. Salmon, *Statistical Explanation and Its Models*, pp. 24-47
M. Scriven, "Causation as Explanation," *Nous,* 1975, pp. 3-10
J. Heckman, "The Scientific Model of Causality," *Sociological Methodology*, 2005, pp. 1-98
P. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986, pp. 945-960

J. Fearon, "Counterfactuals and Hypothesis Testing in Political Science," *World Politics*, 1991, pp. 169-195
G. King, R. Keohane, and S. Verba, *Designing Social Inquiry*, ch. 3
J. Mackie, "Causes and Conditions," *American Philosophical Quarterly*, 1965, pp. 245-264
C. Ragin, *Fuzzy-Set Social Science*, chs. 4, 5
B. Braumoeller, "Causal Complexity and the Study of Politics," *Political Analysis*, 2003, pp. 209-233
B. Braumoeller and G. Goertz, "The Methodology of Necessary Conditions," *American Journal of Political Science*, 2000, pp. 844-858
A. Bennett and C. Elman, ""Complex Causal Relationships and Case Study Methods: The Example of Path Dependence," *Political Analysis*, 2006, pp. 250-267
S. Ashworth, J. Clinton, A. Meirowitz, and K. Ramsey, "Design, Inference, and the Strategic Logic of Suicide Terrorism," *American Political Science Review*, 2008, pp. 269-274
R. Pape, "Methods and Findings in the Study of Suicide Terrorism," *American Political Science Review*, 2008, pp. 275-277
T. Yamamoto, ""Understanding the Past: Statistical Analysis of Causal Attribution," *American Journal of Political Science*, 2012, pp. 237-256
M. Blackwell, "A Framework for Dynamic Causal Inference in Political Science," *American Journal of Political Science*, 2012, pp. 1-17

Threats and Mechanisms

Threats and mechanisms play opposite roles in causal analysis. Threats, or alternative explanations, can undermine causal claims, and thus researchers want to eliminate as many plausible alternative explanations as they can. In contrast, when researchers demonstrate the mechanism(s) by which X influences Y, they strengthen their causal claims considerably. In the past two or three years, social scientists have begun to emphasize the importance of demonstrating causal mechanisms in quantitative research.

In chapter 2 of their book, Shadish, Cook, and Campbell set forth the classic threats to showing true cause and effect. Angrist and Pischke, in their chapter 2, focus on a particularly pervasive threat: heterogeneous selection into or out of the treatment. Gaines and Kuklinski explore four varieties of heterogeneity that can complicate if not derail interpretations of estimated treatment effects. Although their discussion is within the context of experimental design, their cautions apply more generally.

Angrist's *Economic Journal* article is a classic treatment of heterogeneity, including its causes and potential consequences. Gaines and Kuklinski

discuss four sources of heterogeneity within the context of experiments, although their list applies to observational studies as well. Brand and Simon-Thomas present an excellent overview of causal effect heterogeneity. The Roy and Borjas papers explicate and then use what has become known as the "Roy model of heterogeneity. The remaining articles discuss heterogeneity from a variety of disciplines, including history, education, economics, and politics and the life sciences.

(For those seeking to pursue the topic further, the following eight articles offer approaches to reducing the biasing effects of heterogeneity on the estimated parameters of interest. We will not pursue the endeavor in this course.)

The remaining readings address the topic of mediators, also known as mechanisms. Hedstrom and Swedberg adopt an abstract notion that can be represented by simple models, while the others think of mediation in empirical terms. Baron and Kenney is the classic work. Bullock et al. raise the possibility that social scientists will find it very hard to identify and analyze mechanisms, especially in experimental research. Glynn underlines this prediction. The Imai et al. article, which uses Brader et al.'s previously published article on reactions to immigrants as an illustration; Imai et al. will likely shape future thinking about how to incorporate mechanisms into analysis, if they in fact can be incorporated. Verhulst et al. test the thesis that personality traits mediate between genetic factors and political attitudes, and conclude that they do not. Are their analyses and conclusions convincing? Why or why not? Blattman tries to identify mechanisms in his research on war and participation. Does he succeed, at least in terms of Imai et al.'s prescription? How effective is Snow (as described in Freedman) in identifying mechanisms? Does Snow's investigation of cholera offer contemporary researchers any tips on how to proceed in their research?

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch. 2
J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 2
B. Gaines and J. Kuklinski, "Treatment Effects," in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds. *Cambridge Handbook of Experimental Political Science*
J. Brand and J. Simon-Thomas, "Causal Effect Heterogeneity," in S. Morgan, ed., *Hanbook of Causal Analysis for Social Research*, forthcoming
A. Roy, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers (New Series),* 1951, pp.135-146.

G. Borjas, "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 1987, pp. 531-553
S. Wegge, "Occupational Self-Selection of European Immigrants: Evidence from Nineteenth-Century Heese-Cassel," *European Review of Economic History*, 2002, pp. 365-394
S. Morgan, "Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning," *Sociology of Education*, 2001, pp. 341-374
R. Abramitzky, L. Boustan, and K. Eriksson, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 2012, pp. 1832-1856 (also an example of matching, which we discuss later)
E. White, "Self-Selection and Social Life: The Neuropolitics of Alienation—The Trapped and Overwhelmed," *Politics and the Life Sciences*, 1989, pp. 154-169

J. Park, "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models," *American Journal of Political Science*, 2012, pp. 1040-1054
F. Elwert and C. Winship, "Effect Heterogeneity and Bias in Main-Effects-Only Regression Models," in in R. Dechter, H. Geffner, and J. Halpern, eds., *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, ch. 19
M. Humphreys, "Bounds on Least Squares Estimates of Causal Effects in the Presence of Heterogeneous Assignment Probabilities," 2009, unpublished paper
S. Morgan and J. Todd, "A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects, *Sociological Methodology*, 2008, pp. 231-281
D. Green and H. Kern, "Modeling Heterogeneous Treatment Effects in Large-Scale Experiments Using Bayesian Additive Regression Trees," 2010, Unpublished paper
M. Bitler, J. Gelbach, and H. Hoynes, "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment," 2010, Unpublished paper
A. Feller and C. Holmes, "Beyond Toplines: Heterogeneous Treatment Effects in Randomized Experiments," 2009, Unpublished paper
K. Imai and A. Strauss, "Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of Get-Out-the-Vote Campaign, *Political Analysis*, 2011, pp. 1-19

P. Hedstrom and R. Swedberg, eds., *Social Mechanisms*, introduction
D. Waldner, "What are Mechanisms and What Are They Good for?" 2011, Unpublished paper

R. Baron and D. Kenney, "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 1986, pp. 1173-1182

J. Bullock, D. Green, and S. Ha, "Yes, But What's the Mechanism? (Don't Expect an Easy Answer)," *Journal of Personality and Social Psychology*, 2010, pp. 550-558

T. Brader, N. Valentino, and E. Suhay, "What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration," *American Journal of Political Science*, 2008, pp. 959-978

K. Imai, L. Keele, D. Tingley, and T. Yamamoto, "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies," *American Political Science Review*, 2011, pp. 765-789

A. Glynn, "The Product and Difference Fallacies for Indirect Effects," *American Journal of Political Science*, 2012, pp. 257-269

B. Verhulst, L. Eaves, and P. Hatemi, "Correlation not Causation: The Relationship between Personality Traits and Political Ideologies," *American Journal of Political Science*, 2011, pp. 1-18

C. Blattman, "From Violence to Voting: War and Political Participation in Uganda," *American Political Science Review*, 2009, pp. 231-247

D. Freedman, "Statistical Models and Shoe Leather," *Sociological Methodology*, 1991, pp. 291-313

Analysis of Observational Data: Selection on Observables

The term "selection on observables," which is prominent in the econometric literature, is only now becoming a part of political science language. The term refers to an assumption that all of the variables that affect both the outcome and selection into the treatment are properly incorporated into the statistical analysis. This is one of the most crucial assumptions that users of observational data routinely make, and it is reflected in the commonly stated words, "I have controlled for all confounding factors." If the assumption is wrong, the consequence will be the generation of biased estimates of the treatment effect.

Shortcomings of Traditional Workhorse Regression

Until recently, political scientists who used regression analysis reduced their primary task to properly specifying their models, which meant including the right controls and excluding the wrong ones. King, Keohane, and Verba, chapters 4 and 5, effectively summarize this view. Angrist and Pischke present a more formal summary, and Angrist and Krueger present a comprehensive review of empirical strategies, including their strengths and shortcomings. Within a few short chapters, Lieberson

shows the difficulties associated with proper specification. In particular, he shows how the inclusion of control variables can actually increase the bias of an estimated coefficient.

A related problem, which only recently has come to the fore, is that workhouse regression estimates only the average treatment effect. The emphasis on average effect in political science goes back at least to Prezworski and Teune, who were responding to qualitative case studies. Elwert and Winship show that not accounting for heterogeneous treatment effects in main-effects-only regression models will bias estimates, while Rhodes goes further and argues that average treatment effect has no meaningful interpretation when heterogeneous effects exist. Achen proposes simplicity as the best means to address heterogeneity. Do you agree with his prescription? Glasgow et al. undertake a methodologically sophisticated study that they say accounts for heterogeneity. Do they succeed? Would Achen endorse their methodology?

G. King, R. Keohane, and S. Verba, *Designing Social* Inquiry, chs. 4, 5
J. Angrist and J-S. Pischke, ch. 3
J. Angrist and A. Krueger, "Empirical Stategies in Labor Economics," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. 3
S. Lieberson, *Making It Count*, chs. 2-4
A. Przeworski and H. Teune, *Logic of Comparative Social Inquiry*, chs. 5-6
F. Elwert and C. Winship, "Effect Heterogeneity and Bias in Main-Effects-Only Regression Models," in R. Dechter, H. Geffner, and J. Halpern, eds., *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, ch. 19
W. Rhodes, "Heterogeneous Treatment Effects: What Does a Regression Estimate?" *Evaluation Review,* 2010, pp. 334-361
C. Achen, "An Agenda for the New Political Methodology: Microfoundations and ART," *Annual Review of Political Science*, 2002, pp. 423-451
G. Glasgow, M. Golder, and S. Golder, "Who 'Wins'? Determining the Party of the Prime Minister," *American Journal of Political Science*, 2011, pp. 936-953

Heckman Two-Step Model and Propensity Scores and Matching

Above, we explored the shortcomings associated with the use of control variables in regression models. Keep those shortcomings in mind as you read through the following articles.

The first two articles focus on Heckman's two-step model, which is designed to account for self-selection. Heckman sets forth the logic of his model, which political scientists have only recently begun to use. Von Stein's article continues as a leading application of Heckman in the discipline. Lemke and Reed is another good illustration. What should one make of LaLonde's demonstration that the Heckman two-step model does not necessarily generate the right estimates?

The sections from Angrist and Pischke and the following four articles set forth the potential outcomes-propensity scores logic. Rosenbaum and Rubin have been among the most visible advocates of propensity scores. A very important task is to understand exactly how the use of propensity scores differs from the use of ordinary regression analysis. In the end, do they really differ, or do the same problems beset both approaches?

Gilligan and Sergenti, Nielsen et al., Humphreys and Weinstein, and Xie et al. all use matching methods to identify cause and effect. Do they succeed? What do you conclude from the Kam and Palmer versus Henderson and Chatfield exchange? What should one make of Arceneaux et al.'s demonstration that matching does not necessarily produce the right estimates?

J. Heckman, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 1976, pp. 475-492
J. Von Stein, "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance," *American Political Science Review*, 2005, pp. 611-62
D. Lemke and W. Reed, "War and Rivalry among Great Powers," *American Journal of Political Science*, 2001, pp. 457-469
R. LaLonde, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 1986, pp. 604-620
J. Angrist and P-S. Picshke, *Mostly Harmless Econometrics*, section 3.3
P. Rosenbaum and D. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, pp. 41-55
P. Rosenbaum and D. Rubin, "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 1985, pp. 33-38
P. Rosenbaum and D. Rubin, "Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score," *Journal of the American Statistical Association*, 1984, pp. 516-524

J. Sekhon, "The Neyman-Rubin Model of Causal Inference and Estimation Via Matching Methods," in J. Box-Steffensmeier, H. Brady, and D. Collier, eds., *The Oxford Handbook of Political Methodology*
M. Gilligan and E. Sergenti, "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference," *Quarterly Journal of Political Science*, 2008, pp. 89-122
R. Nielsen, M. Findley, Z. Davis, T. Candland, and D. Nielson, "Foreign Aid Shocks as a Cause of Violent Armed Behavior," *American Journal of Political Science*, 2011, pp. 219-232
M. Humphreys and J. Weinstein, "Demobilization and Reintegration," *Journal of Conflict Resolution*, 2007, pp. 531-567
Y. Xie, J. Brand, and B. Jann, ""Estimating Heterogeneous Treatment Effects with Observational Data," 2011, Unpublished paper
C. Kam and C. Palmer, "Reconsidering the Effects of Education on Political Participation," *Journal of Politics*, 2008, pp. 612-631
J. Henderson and S. Chatfield, "Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation," *Journal of Politics*, 2011, pp. 646-658
C. Kam and C. Palmer, "Rejoinder: Reinvestigating the Causal Relationship between Higher Education and Political Participation," *Journal of Politics*, 2011, pp. 659-663
K. Arceneaux, A. Gerber, and D. Green, "A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Empirical Benchmark," *Sociological Methods and Research*, 2010, pp. 256-282
R. Abramitzky, L. Boustan, and K. Eriksson, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 2012, pp. 1832-1856 (also an example of matching, which we discuss later)

Counterfactual Decomposition and Quantile Regression

While both counterfactual decomposition and quantile regression select on observables only, both try to address heterogeneity, albeit from very different directions. The basic idea of the Blinder-Oaxaca decomposition technique, named after two economists, began in labor economics, where researchers often want to know if differences in wages between two groups—men and women, say—arise because of differences in qualifications and credentials when both groups receive the same treatment (the explained component) or because one group is more favorably treated than the others given the same individual characteristics (the unexplained component). The method has been widely used to study racial and gender wage differentials and discrimination.

Quantile regression—the leading expert on the method is UI economist Roger Koenker—emerged out of the recognition that a treatment can do more than change the location of a conditional mean location. Most important, a treatment can affect the shape of the distribution. For example, the recent housing crisis clearly brought down housing prices, overall. A remaining question is whether the crisis affected housing prices at the high end more than prices at the low end (or vice-versa). Whereas OLS relies on means, quantile regression uses medians. Not only are decomposition and quantile regression designed to address heterogeneity, their use opens the door to a wholly new set of questions.

The two original articles on decomposition are Blinder and Oaxaca, which happened to be published in the same year. Jann provides a brief overview of decomposition from a Stata perspective. Kim and Reed and Chiba use decomposition to address substantive questions, one in the area of race, the other in the area of military conflict.

Koenker and Hallock's essay on quantile regression serves as a useful introduction to the technique and its purposes. Angrist and Pischke's ch. 7 provides an additionally useful overview. Breunig applies quantile regression to the study of the effects of changes within national budgets.

A. Blinder, "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 1973, pp. 436-455
R. Oaxaca, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 1973, pp. 693-709
B. Jann, "The Blinder-Oaxaca Decomposition for Linear Regression Models," *The Stata Journal*, 2008, pp. 453-479
C. Kim, "Decomposing the Change in the Wage Gap between White and Black Men over Time, 1980-2005: An Extension of the Blinder-Oaxaca Decomposition Method," *Sociological Methods and Research,* 2010, pp. 619-651
W. Reed and D. Chiba, "Decomposing the Relationship between Contiguity and Militarized Conflict," *American Journal of Political Science*, 2010, pp. 61-73
R. Koenker and K. Hallock, "Quantile Regression," *Journal of Economic Perspectives*, 2001, pp. 143-156
J. Angrist and J-S Pischke, *Mostly Harmless Econometrics*, ch. 7
C. Breunig, "Reduction, Stasis, and Expansion of Budgets in Advanced Democracies," *Comparative Political Studies, 2011, pp. 1068-1088*

Analysis of Observational Data: Selection on Un-observables

Although the approaches enumerated below have been around in one form or another for some time, economists have done everyone a favor by categorizing them all under the term, "selection on un-observables." Unlike the approaches in the preceding section, those that follow immediately below attempt to eliminate biased estimates by making the objects of comparison equivalent via imaginative design techniques. Specifically, we review the use of instrumental variables, differences-in-differences designs, and regression discontinuity designs.

Instrumental Variables

Instrumental variables and two-stage least squares, the most common way to implement instruments, have been around for a long time. Only recently have social scientists begun fully to understand their value and their limitations. Angrist has been an intellectual leader in IV. The first four readings, below, set forth the essential logic of IV, and the following three derive implications of treatment heterogeneity for the use of IV (most of the implications are not good news). The next five readings are examples of IV in use to study substantive questions. Angrist brought the substantive value of IV to the fore in his study of the Vietnam era draft lottery's effect on lifetime earnings. The Miguel et al. study has strongly influenced the study of civil conflict.

J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 4
J .Angrist, G. Imbens, and D. Rubin, ""Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, pp. 444-455
J. Angrist and A. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 2001, pp. 69-85
A. Sovey and D. Green, "Instrumental Variables Estimation in Political Science: A Readers' Guide," *American Journal of Political Science,* 2009, pp. 188-200
J. Angrist, "Treatment Effect Heterogeneity in Theory and Practice," *The Economic Journal*, 2004, pp. C52-C84
A. Manning, "Instrumental Variables for Binary Treatments with Heterogeneous Treatment Effects: A Simple Exposition," *Contributions to Economic Analysis*, 2004, pp. 1-14
A. Basu, J. Heckman, S. Navarro-Lozano, and S. Urzua, "Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: An Application to Treatments of Breast Cancer Patients," *Health Economics*, 2007, pp. 1133-1157
L. Bartels, "Instrumental and 'Quasi-Instrumental' Variables," *American Journal of Political Science*, 1991, pp. 777-800

J. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," *American Economic Review,* 1990, pp. 313-336
J. Angrist, "Instrumental Variables Methods in Experimental Criminological Research: What, Why, and How," *Journal of Experimental Criminology*, 2006, pp. 23-44
E. Miguel, S. Satyanath, and E. Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, pp. 725-753
B. Savun and D. Tirone, "Foreign Aid, Democratization, and Civil Conflict: How Does Democracy Aid Affect Civil Conflict?" *American Journal of Political Science*, 2011, pp. 233-246

Fixed Effects Panel Data and Differences-in-Differences

Both fixed effects and differences-in-differences estimates implicitly assume that omitted variables are time- and/or group-invariant. The former approach uses dummy variables to designate time, the individual units, or both as a way to control for unobserved factors. The two problems are, first, that the equations normally include a ton on variables, and, second, that the dummy variables are not specific in what they control. They throw out the baby with the bathwater. Differences-in-differences estimates depend on two assumptions: first, much can be learned about cause and effect by examining a phenomenon of interest both before and after a real-world intervention to determine the change; and it is crucial to include, for purposes of comparison, units in a control condition that are as similar as possible to those in the treated condition. Thus comes the term "difference-in-differences." Economists' primary contributions have been to improve the rigor of such studies and to show creativity in their searches for relevant data. Note that differences-in-differences, like fixed effects models, require across-time data.

The first three readings are an excellent discussion of some of the problems that attend panel data, especially when those data, as is common in international relations, consist of dyads. Notice that some very prominent methodologists do not offer compelling solutions to the problems they identify.

Differences-in-differences estimation resembles experimental research. However, randomization typically is absent in the former. Donald Campbell is generally viewed as the creator of quasi-experimental designs, a term that predated differences-in-differences, and his "Reforms as Experiments" represents his earliest and still-impressive thinking about them. It is useful to read chapters 4, 5, and 6 in Shadish et al.; at a

minimum, read chapter 6. As you read these chapters, you might conclude that you've been assigned a football team's playbook. There are a lot of X's and O's. In fact, understanding the strengths and weaknesses of these designs will be one of the most important things you do in this course. In light of these chapters, how compelling are the Lewis-Beck and Alford and Campbell and Ross studies? The Pelowski article reports a quasi-experimental study that was completed in 1971! At the time, the Northwestern University Department of Political Science, Pelowski's home department, included some of the leading international relations scholars in the world.

The next six readings focus on the same methodology, albeit in economists' language. The Angrist and Pischke chapter and article, along with the Rubin and Rosenbaum articles, make the case for differences-in-differences designs. Leamer, Bertrand et al., and Donald and Lange express less enthusiasm. The remaining readings report substantive studies that use differences-in-differences designs. The Adabie et al. paper is an excellent example of how differences-in-differences logic can be applied to case studies.

D. Green, S. Kim, and D. Yoon, "Dirty Pool," *International Organization*, 2001, pp. 441-468
J. Oneal and B. Russett, "Clear and Clean: The Fixed Effects of the Liberal Peace," *International Organization,* 2001, pp. 469-485
G. King, "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data," *International Organization*, 2001, pp. 497-50

D. Campbell, "Reforms as Experiments," *American Psychologist*, 1969, pp. 409-429
W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, chs. 2, 4, 5, 6, 7
M. Lewis-Beck and J. Alford, "Can Government Regulate Safety? The Coal Mine Example," *American Political Science Review*, 1980, pp. 745-756
D. Campbell and H. Ross, "The Connecticut Crackdown on Speeding," in E. Tufte, ed., *The Quantitative Analysis of Social Problems*
A. Pelowski, "On the Use of a Quasi-Experimental Design in the Study of International Organization and War," *Journal of Peace Research*, 1971, pp. 279-285
J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch. 5
J. Angrist and J-S Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 2010, pp. 3-30

D. Rubin, "For Objective Causal Inference, Design Trumps Analysis," *The Annals of Applied Statistics*, 2008, pp. 808-840
D. Rosenbaum, "Choice as an Alternative to Control in Observational Studies," *Statistical Science*, 1999, pp. 259-278
E. Leamer, "Tantalus on the Road to Asymptopia," *Journal of Economic Perspectives*, 2010, pp. 31-46
M. Bertrand, E. Duflo, and S. Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 2004, pp. 249-275
S. Donald and K. Lang, "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics*, 2007, pp. 221-233
D. Card, "The Impact of the Mariel Boatlife on the Miami Labor Market," *Industrial and Labor Relations Review*, 1990, pp. 245-257
D. Card and A. Krueger, "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 1994, pp. 772-784
J. Lyall, "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya," *Journal of Conflict Resolution*, 2009, pp. 331-362
S. Anzia and C. Berry, "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science*, 2011, pp. 478-493
W. Bullock and J. Clinton, "More a Molehill than a Mountain: The Effects of the Blanket Primary on Electoral Officials' Behavior from California," *Journal of Politics*, 2011, pp. 915-930
A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating The Effect of California's Tobacco Control Program," Unpublished paper

Regression Discontinuity

In the words of Angrist and Pischke (*Mostly Harmless Econometrics*, p. 251), "Regression discontinuity (RD) designs exploit precise knowledge of the rules determining treatment. RD identification is based on the idea that in a highly rule-based world, some rules are arbitrary and therefore provide good experiments." In other words, RD designs attempt to use selection to advantage, rather than eliminate its effects.

The first four readings provide useful overviews of RD designs. Psychologists authored the first two, economists the last two. Do they share a common perspective, or do the two groups differ in perspectives? The next five readings report substantive findings using RD, and the final reading compares the results of RD with those derived from experiments.

What do the authors of the final reading conclude? What should one make of this conclusion?

W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch.7
T. Cook, "Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics," *Journal of Econometrics*, 2008, pp. 636-654
J. Angrist and J-S. Pischke, *Mostly Harmless Econometrics*, ch.7
G. Imbens and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2007, pp. 615-635
D. Lee, E. Moretti, and M. Butler, "Do Voters Affect or Elect Policies: Evidence from the U.S. House," *Quarterly Journal of Economics*, 2004, pp. 807-859
D. Lee, "Randomized Experiments from Non-Random Selection in U.S. House Elections," *American Political Science Review*, 2008, pp. 675-697
A. Eggers and J. Hainmueller, "MPs for Sale? Returns to Office in Postwar British Politics," *American Political Science Review*, 2009, pp. 1-21
E. Gerber and D. Hopkins, "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy," *American Journal of Political Science*, 2011, pp. 326-339
A. Gerber, D. Kessler, and M. Meredith, "The Persuasive Effects of Direct Mail: A Regression Discontinuity Design," *Journal of Politics*, 2011, pp. 140-155
D. Green, T. Leong, H. Kern, A. Gerber, and C. Larimer, ""Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks," *Political Analysis*, 2009, pp. 400-417

The Generation and Interpretation of Experimental Data

Experimental research pervades nearly every mature scientific discipline. Physics, chemistry, astronomy, and psychology all began as non-experimental endeavors. In every instance, early practitioners proclaimed that their discipline could never be experimental; in every instance they were wrong. During the past 15 years, political scientists have come to embrace experimental research. It would be grossly premature to suggest that the discipline has become an experimental discipline.

Scientists view random assignment experiments as the gold standard for determining cause and effect because internal validity is high. On the other hand, critics point to the lack of external validity and claim that the results of experiments cannot be generalized. In other words, the primary strength of observational data, generalization, purportedly is the primary weakness of

experimental data, while the primary strength of experimental data, internal validity, is the primary weakness of observational data. Thus, says conventional wisdom, the choice of one type of data over the other entails a tradeoff. Although true, conventional wisdom simplifies.

The following readings fall into two (large) groupings. The first grouping provides an overview of experimental research in political science. The included readings illustrate the types and uses of experiments. In terms of types, one can distinguish among field, survey, and laboratory experiments; between one-shot and longitudinal experiments; between within- and between-subjects experiments; and between experiments that use dosages and those that do not. In terms of purpose, political scientists design experiments to test and modify formal theory, to make inferences about the non-experimental world, and to determine how programs and policies work in real-world settings. The second grouping emphasizes challenges in experimental research that political scientists are only now beginning to recognize and address. These challenges go well beyond internal and external validity, terms that seem no longer as relevant and useful as they once were. The potentially the most crucial challenge of all: finding ways to overcome violations of the so-called Stable Unit Treatment Value Assumption (yup, that is the correct term!). Although included in the section on experiments, violations of SUTVA apply to observational studies as well. Several readings in this third subsection address observational studies.

Types and Purposes of Experimental Designs

Druckman et al. document the explosion in experimental studies. Fisher, Shadish et al., Druckman et al., and Green and Gerber set forth the essential logic, in simple, and as we shall see, possibly simplistic terms. Dickson summarizes the differences between economics-based and psychology-based experiments. The differences are stark! The remaining readings give examples of the different types and uses of experiments. Note that students of comparative politics and international relations are now routinely completing experimental studies. Students of comparative politics, in particular, are generating some of the most exciting experimental research.

J. Druckman, D. Green, J. Kuklinski, and A. Lupia, "The Growth and Development of Experimental Science in the *American Political Science Review*, 2006, pp. 627-636
R. Fisher, *Design of Experiments*, chs. 1-3
W. Shadish, T. Cook and D. Campbell, *Experimental and Quasi-Experimental Designs*, ch. 8
J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds., *Cambridge Handbook of Experimental Political Science*, chs. 1-2

D. Green and A. Gerber, “The Downstream Benefits of Experimentation,” *Political Analysis*, 2002, pp. 394-402
E. Dickson, “Economics versus Psychology Experiments,” in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds., *Cambridge Handbook of Experimental Political Science*
A. Gerber, D. Green, and C. Larimer, “Social Pressure and Vote Turnout: Evidence from A Large-Scale Field Experiment,” *American Political Science Review*, 2008, pp. 33-48
R. Dawes, J. Orbell, R. Simmons, and J. Van De Kragt, “Organizing Groups for Collective Action,” *American Political Science Review*, 1986, pp. 1171-1185
E. Gerber, R. Morton, and T. Rietz, “Minority Representation in Minority Districts,” *American Political Science Review*, 1998, pp. 127-144
M. Peffley and J. Hurwitz, “Persuasion and Resistance: Race and the Death Penalty in America,” *American Journal of Political Science*, 2007, pp. 996-1012
E. Ostrom, J. Walker, and R. Gardner, “Covenants with and without a Sword,” *American Political Science Review*, 1992, pp. 404-417
K. Arceneaux and and R. Kolodny, “Educating the Least Informed: Group Endorsements in a Grassroots Campaign,” *American Journal of Political Science*, 2009, pp. 755-770
G. Quattrone and A. Tversky, “Contrasting Rational and Psychological Analyses of Political Choice,” *American Political Science* Review, 1988, pp. 719-736
P. Tetlock and R. Lebow, “Pocking Counterfactual Holes in Covering Laws,” *American Political Science Review,* 2001, pp. 829-843
R. Herrmann, P. Tetlock, and P. Visser, “Mass Public Decisions to Go to War,” *American Political Science Review*, 1999, pp. 553-573
S. Iyengar, “Television News and Citizens’ Explanations of National Issues,” *American Political Science Review*, 1987, pp. 815-832
N. Geva, J. Mayhar, and J. Skorick, “The Cognitive Calculus of Foreign Policy Decision Making: An Experimental Assessment,” *Journal of Conflict Resolution*, 2000, pp. 447-471
P. Sniderman and T. Piazza, *The Scar of Race*, pp. 102-104
J. Habyarimana, M. Humphreys, D. Posner, and J. Weinstein, “Why Does Ethnic Diversity Undermine Public Goods Provision?” *American Political Science Review*, 2007, pp. 709-726
M. Fiorina and C. Plott, “Committee Decisions under Majority Rule: An Experimental Study,” in D. Kinder and T. Palfrey, eds., *Experimental Foundations of Political Science*

A. Posner, "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi," *American Political Science Review*, 2004, pp. 529-545
M. Lodge, C. Taber, and B. Verhulst, "Conscious and Unconscious Information Processing with Implications for Political Science," in J. Druckman, D. Green, J. Kuklinski, and A. Lupia, eds., *Handbook of Experimental Political Science*
A. Coleman and E. Ostrom, "Experimental Contributions to Collective Action Theory," in J. Druckman et al., eds., *Handbook of Experimental Political Science*
M. Findley, D. Nielson, and J. Sharman, "The Causes of Compliance in International Relations: Evidence from a Field Experiment on Financial Transparency," 2011, Unpublished paper
D. Green, P. Aronow, D. Bergan, P. Greene, C. Paris, and B. Weinberger, ""Does Knowledge of Constitutional Principles Increase Support for Civil Liberties? Results from a Randomized Field Experiment," *Journal of Politics*, 2011, pp. 463-476
D. Chong and J. Druckman, ""Dynamic Public Opinion: Communication Effects over Time," *American Political Science Review*, 2010, pp. 663-680
A. Gerber, G. Huber, and E. Washington, "Party Affiliation, Partisanship, and Political Beliefs," *American Political Science Review*, 2010, pp. 720-744
D. Mitchell, "It's About Time: The Lifespan of Information in a Multi-week Campaign," *American Journal of Political Science*, 2011, pp. 298-311
G. Grossman and D. Baldassari, "The Impact of Elections on Cooperation: Evidence from a Lab-in-Field Experiment in Uganda," *American Journal of Political Science*, 2012, pp. 964-985

Challenges and Responses to Them

Not too long in the past, critics of experimental research focused almost exclusively on the lack of external validity, the capacity to speak beyond the specific experiment. While few would suggest that the external-versus-internal validity debate is now irrelevant, it arguably has taken a back seat to other challenges.

The first six articles address external validity. Druckman and Kam is one of the latest words on this matter, and they argue (convincingly?) that the use of students is not a problem for external validity. Mook's old article argues that the preoccupation with external validity is misplaced, and reviews two classic studies to make his point. In contrast, Barabas and Jerit apply the external validity criterion in a convincing way.

The next two studies address the potential problem of pretreatment: subjects who enter an experimental setting have already been exposed to the treatment of interest to the experimenter. Then the question becomes, how does one interpret the results? The first assigned reading raises the possibility that pretreatment can be a problem, the second provides an instance where it is a problem.

An equally important concern, especially for users of survey and laboratory experiments, is whether the "captive audience" aspect of survey and laboratory experiments can bias the estimate of the effect of interest. Gaines and Kuklinski argue for the incorporation of selection into the random assignment experiment. Orbell and Dawes and Druckman, Fein, and Leeper show that allowing subjects to select in or out of the experimental treatment strongly affects the outcomes.

The next five readings illustrate and address a problem that afflicts nearly all field experiments: noncompliance. Whenever some individuals who are assigned to the treatment group never receive the treatment, random assignment no longer holds. Consequently, the experimenter finds him- or herself facing the same problems that a user of observational studies faces. Hansen and Bowers and Horiuchi suggest ways to overcome the noncompliance problem.

Yet another challenge arises when one thinks in terms of levels of analysis higher than the individual unit. Then a particularly interesting question arises: At what level(s) of analysis does the experimenter randomize? The two pieces by Betsy Sinclair and, in one case, coauthors offer some useful guidelines.

Finally, most social scientists assume autonomous individual actors (individuals, states, nations, and so forth) whose decisions to receive or not receive treatments are not affected by the actions of others. This assumption, of course, is blatantly wrong. Strategic interactions are the defining feature of political elites. Moreover, most people are embedded in social networks, where considerable influence undoubtedly occurs. To the extent that the no-interference assumption is wrong, estimates of treatment effects will likely be biased. To put the matter more bluntly, if social interactions moderate treatment effects, then one can justifiably eye much past research as worthy of re-visitation. Even data collected in random sample surveys are not completely immune to possible contamination.

Currently, one of the most active areas of research across all the social sciences is to find ways to estimate treatment effects given interference arising from social interactions. The enormity of this challenge can hardly be overstated; and the literature is almost overwhelming. Indeed, trying to organize this literature proves difficult. Included in the readings is Manski's influential work on what he calls the reflection problem, as well as other economists' research on the interference problem. Fowler approaches the interference problem directly, in terms of social networks, while Simmons and Elkins turn a social-interaction-based process, policy diffusion, into a phenomenon to be studies in its own right. The remaining four articles, including a sophisticated analysis by UI authors Bowers and Fredrickson, approach the interference problem in especially sophisticated ways. A question for IR students: In light of emerging work, what should political scientists make of the use of dyads as the unit of analysis? Is there a reason to be concerned?

L. Berkowitz and E. Donnerstein, "External Validity is more than Skin Deep: Some Answers to Criticisms of Laboratory Experiments," *American Psychologist*, 1982: 245-257
D. Sears, "College Sophomores in the Laboratory: Influences of a Narrow Database on Social-Psychological Views of Human Nature," *Journal of Personality and Social Psychology*, 1986, pp. 515-530
B. Anderson, "External Validity of 'Trivial' Experiments: The Case of Laboratory Aggression," *Journal of General Psychology*, 1997, pp. 19-41
J. Druckman and C. Kam, "Students as Experimental Participants: A Defense of the 'Narrow Data Base,'" in Druckman et al., eds., *Handbook of Experimental Political Science*
D. Mook, "In Defense of External Validity," *American Psychologist*, 1983, pp. 379-387
J. Barabas and J. Jerit, "Are Survey Experiments Externally Valid?" *American Political Science Review*, 2010, pp. 226-242

B. Gaines, J. Kuklinski, and P. Quirk, "The Logic of the Survey Experiment Reexamined," *Political Analysis*, 2007, pp. 1-20
J. Druckman and T. Leeper, "Learning More from Political Communication Experiments:
Pretreatment and Its Effects, *American Journal of Political Science*, 2012, pp. 875-896

B. Gaines and J. Kuklinski, "Experimental Estimation of Heterogeneous Treatment Effects
Related to Self-Selection," *American Journal of Political Science*, 2011, pp. 724-736

J. Orbell and R. Dawes, "A 'Cognitive Mister' Theory of Cooperators' Advantage,"
American Political Science Review, 1991*,* pp. 515-528
J. Druckman, J. Fein, and T. Leeper, "A Source of Bias in Public Opinion Stability," *American*
Political Science Review, 2012, pp. 430-454

A. Gerber and D. Green, "The Effects of Canvassing, Direct Mail, and Telephone
Contact on Voter Turnout: A Field Experiment," *American Political Science Review*, 2000, pp. 653-663
K. Imai, "Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical
Methods for Field Experiments," *American Political Science Review*, 2005, pp. 283-300
A. Gerber and D. Green, "Correction to Gerber and Green (2000), Replication of
Disputed Findings, and Reply to Imai (2005)," *American Political Science Review*, 2005, pp. 301-313
B. Hansen and J. Bowers, "Attributing Effects to a Cluster Randomized Get-Out-The-
Vote Campaign," *Journal of the American Statistical Association*, 2009, pp. 873-885
B.Hansen and J. Bowers, "Covariate Balance in Simple, Stratified and Clustered Comparative
Studies," *Statistical Science*, 2008, pp. 219-236
Y. Horiuchi, K. Imai, and N. Taniguchi, "Designing and Analyzing Randomized Experiments:
Application to a Japanese Election Survey Experiment," *American Political Science Review*, 2007, pp. 669-687

B. Sinclair, "Design and Analysis of Experiments in Multi-Level Populations," in Druckman et
al., eds., *Handbook of Experimental Political Science*
B. Sinclair, M. McConnell, and D. Green, "Detecting Spillover Effects: Design and Analysis of
Multi-Level Experiments," *American Journal of Political* Science, 2012, pp. 1055-1069

C. Manski, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review*
of Economic Studies, 1993, pp. 531-542
C. Manski, "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*,

2000, pp. 115-136
C. Manski, "Identification of Treatment Response with Social Interactions," *The Econometrics Journal*, forthcoming
M. Sobel, What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 2006, pp. 1398-1407
M. Hudgens and M. Halloran, "Toward Causal Inference with Interference," *Journal of the American Statistical Association*, 2008, pp. 832-842
E. Duflo and E. Saez, "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 2003, pp. 815-842
S. Currarini, M. Jackson, and P. Pin, "Identifying the Roles of Race-based Choice and Chance in High School Friendship Network Formation," *Proceedings of the National Academy of Sciences*, 2010, pp. 4857-4861
S. Currarini, M. Jackson, and P. Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 2009, pp. 1003-1045
F. Drago and R. Galbiati, "Indirect Effects of a Policy Altering Criminal Behaviour: Evidence from the Italian Prison Experiment," Unpublished paper
J. Fowler, "Connecting the Congress: A Study of Cosponsor Networks," *Political Analysis*, 2006, pp. 456-487
J. Fowler, J. Rosenquist, and N. Christkis, "Social Network Determinants of Depression," *Molecular Psychiatry*, 2011, pp. 273-281
J. Fowler, C. Apicella, F. Marlowe, and N. Christkis, "Social Networks and Cooperation in Hunter-Gathers," *Nature*, forthcoming
B. Simmons and Z. Elkins, "The Globalization of Liberalization: Policy Diffusion in the International Political Economy," *American Political Science Review*, 2004, pp. 1-19
R. Sampson, "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure," *American Journal of Sociology*, 2008, pp. 189-231
D. Nickerson, "Is Voting Contagious? Evidence from Two Field Experiments," *American Political Science Review*, 2008, pp. 49-58

P. Aronow, "A General Method for Detercting Interference between Units in Randomized
Experiments," 2010, unpublished paper
J. Bowers, M. Fredrickson, and C. Panagopoulos, "Reasoning about Inference between Units: A
General Framework, *Political Analysis*, forthcoming

Coming Full Circle?

A particularly interesting development in political science experimental research is a return to Fisher's concept of randomization inference. It represents a departure from the commonly used classical statistical tests in the frequentist framework. It also differs from most disciplinary empirical research in that the inference is directly to the sample being used. Jake Bowers has been an intellectual leader in this movement.

Randomization inference seemingly values in-depth understanding of a few cases. This raises two questions: First, what is happening to the quest for general statements? Do advocates of randomization inference no longer care about generalization, or do they view generalization is lying in theory, not in the empirical data? Put another way, does randomization inference better represent Popper's view of science than empirical research as political scientists have practiced it for more than half a century? Second, does randomization inference comport with some the arguments that the discipline's leading qualitative researchers have been making? With respect to this second question, Henry Brady and David Collier co-edited *Rethinking Social Inquiry: Diverse Tools, Shared Standards* in 2004. It has been a highly influential volume, and is worthy of every graduate student's time and effort.

Popper and Randomization Inference

K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, ch. 1
R. Fisher, *The Design of Experiments*, all
J. Fowler and C. Kam, "Beyond the Self: Social Identity, Altruism, and Political Participation,"
Journal of Politics, 2007, pp. 813-827
L. Keele, C. McConnaughy, and I. White, "Strengthening the Experimenter's Toolbox:
Statistical Estimation of Internal Validity," *American Journal of Political Science*, 2012, pp. 484-499
J. Bowers and C. Panagopoulos, "Fisher's Randomization Mode of Statistical Inference, Then
and Now," *American Journal of Political Science*, Revise-and-resubmit

The Place of Qualitative Research in a Quantitatively-Oriented Discipline

H. Brady and D. Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, all