Second Edition

POLITICAL SCIENCE RESEARCH IN PRACTICE

Edited by
Akan Malici and Elizabeth S. Smith

POLITICAL SCIENCE RESEARCH IN PRACTICE

Nothing rings truer to those teaching political science research methods: students hate taking this course. Tackle the challenge and turn the standard research methods teaching model on its head with *Political Science Research in Practice.* Akan Malici and Elizabeth S. Smith engage students first with pressing political questions and then demonstrate how a researcher has gone about answering them, walking them through real political science research that contributors have conducted. Through the exemplary use of a comparative case study, field research, interviews, textual and interpretive research, statistical research, survey research, public policy and program evaluation, content analysis, and field experiments, each chapter introduces students to a method of empirical inquiry through a specific topic that will spark their interest and curiosity. Each chapter shows the process of developing a research question, how and why a particular method was used, and the rewards and challenges discovered along the way. Students can better appreciate why we need a science of politics – why methods matter – with these first-hand, issue-based discussions.

The second edition now includes:

- Two completely new chapters: one on field experiments and one on the textual/interpretative method;
- New topics, ranging from the Arab Spring to political torture, from politically sensitive research in China to social networking and voter turnout;
- Revised and updated "Exercises and Discussion Questions" sections;
- Revised and updated "Interested to Know More" and "Recommended Resources" sections.

Akan Malici is Professor of Politics and International Affairs at Furman

University. He authored *When Leaders Learn and When They Don't* (SUNY 2008) and *The Search for a Common European Foreign and Security Policy* (Palgrave 2008). He co-authored *U.S. Presidents and Foreign Policy Mistakes* (Stanford 2011) and *Role Theory and Role Conflict in U.S.–Iran Relations* (Routledge 2016), and co-edited *Re-thinking Foreign Policy Analysis* (Routledge 2011). He teaches classes in International Politics and Research Methodology.

Elizabeth S. Smith is Professor of Politics and International Affairs at Furman University. She received her Ph.D. in American politics with a minor in political psychology from the University of Minnesota. She has taught research methodology for many years. Her work appears in the *Journal of Political Science Education, Polity, Political Psychology, The Handbook for Teaching Social Issues*, the *American Education Research Journal* and in *Competition in Theory and Practice* (Sense Publishers 2009).

"Most methodology books focus on the dry nuts and bolts of empirical inquiry, and often fail to provide an engaging pedagogical context for undergraduates. This book is different. By embedding methodological discussions within the context of important substantive questions, this volume conveys the science of politics in action. I only wish that such a book were available when I was a student."

Howard Lavine, *Arleen C. Carlson Professor of Political Science, University of Minnesota*

"Malici and Smith do two things we rarely see in research methods texts: they cover the rich, broad spectrum of empirical approaches in the discipline and they convey these techniques through first-hand examples. The result is a book that is engaging, accessible, and uniquely valuable to political science students and instructors."

Francis Neely, *Associate Professor, San Francisco State University*

"The timely update of Malici and Smith's book helps students understand how political scientists investigate real-world problems using a full spectrum of methodological tool sets. It continues the previous edition's one-of-a-kind storytelling approach that makes political science research approachable, relevant, and even fun. It will be welcomed by both teachers and students of political science research methods alike."

Yi Edward Yang, *Professor of Political Science, James Madison University*

POLITICAL SCIENCE RESEARCH IN PRACTICE

Second Edition

Edited by

Akan Malici and Elizabeth S. Smith

CONTENTS

ILLUSTRATIONS

FIGURES

TABLES

NOTES ON CONTRIBUTORS

Deven Carlson is Presidential Research Professor, Associate Professor of Political Science, and Associate Director for Education at the National Institute for Risk and Resilience at the University of Oklahoma. He received his B.A. from St. John's University in Collegeville, MN and his M.P.A., M.A. and Ph.D. from the University of Wisconsin-Madison. His work has been published in *Journal of Policy Analysis and Management, Educational Evaluation and Policy Analysis, Journal of Urban Economics, Policy Studies Journal, Sociology of Education*, and many other journals and edited volumes.

Joshua M. Cowen, Ph.D., is Associate Professor of Education Policy and Co-Director of the Education Policy Innovation Collaborative at the College of Education at Michigan State University. His research currently focuses on teacher quality, student and teacher mobility, program evaluation and education policy. His work has been published in the *American Educational Research Journal, Economics of Education Review, Educational Evaluation and Policy Analysis, Education Finance and Policy, Journal of Policy Analysis and Management, Sociology of Education*, and *Teachers College Record*, among other outlets, and has been funded by a diverse array of philanthropies, including the Laura and John Arnold Foundation, the Russell Sage Foundation, the Smith Richardson Foundation, the Spencer Foundation and the Walton Family Foundation. He has worked on program evaluations in Florida, Louisiana, Kentucky, Michigan and Wisconsin, and worked directly with state and district policymakers in Kentucky, Michigan and Wisconsin to inform decision-making. Dr. Cowen is presently co-editor of *Educational Evaluation and Policy Analysis*, the flagship peer-reviewed journal for education policy in the United States.

Zaid Eyadat is a Professor of Political Science and Human Rights, expert on

international and comparative politics and the chairperson of the board of trustees and the advisory board of ARDD (Arab Renascence for Development and Democracy). He is the founding chairperson of the Human Rights Department, which later became the Political Development Department, and the founding Dean of the Prince al-Hussein School of International Studies at the University of Jordan. His training and research interests are in the fields of international politics, comparative politics, international political theory, modeling and game theory. He is a leading and distinguished expert on Middle East politics, human rights, Islamic thought and Islam and human rights. His most recent book is titled *Islam, State, and Modernity: Mohammed Abed al-Jabri and the Future of the Arab World.* (Palgrave Macmillan 2017). His research has also been published in many leading journals.

David J. Fleming is an Associate Professor of Politics and International Affairs and a Senior Researcher at the Riley Institute at Furman University. He received his Ph.D. from the University of Wisconsin-Madison in 2009. Professor Fleming has participated in a number of program evaluations, including studies examining Montessori education, summer programs, school-based health clinics and school vouchers. He teaches classes in American politics, research methods and public policy.

Sarah Allen Gershon is Associate Professor of Political Science at Georgia State University. Her research interests include political communication, gender politics, and race and ethnicity. She has published articles in *Politics, Groups and Identities, Political Communication, Political Research Quarterly, Journal of Politics* and *Social Science Quarterly.* Professor Gershon is co-editor (with Dr. Nadia Brown) of *Distinct Identities: Minority Women in U.S. Politics* (Routledge 2016).

James L. Guth is William R. Kenan, Jr. Professor of Politics and International Affairs at Furman University. He received his B.S. from the University of Wisconsin and Ph.D. from Harvard University. He is the co-author or co-editor of *Religion and the Struggle for European Union, The Oxford Handbook of Religion and American Politics, The Bully Pulpit, Religion and the Culture Wars*, and *The Bible and the Ballot Box.* His work has also appeared in the *American Journal of*

Political Science, Journal of Politics, Political Research Quarterly, Public Opinion Quarterly, European Union Politics, Journal of Common Market Studies, Sociological Analysis, Journal for the Scientific Study of Religion and many other journals and edited collections.

Katherine Palmer Kaup is Professor of Politics and International Affairs and Asian Studies at Furman University. She is author of *Creating the Zhuang: Ethnic Politics in China* (Lynne Rienner Press 2000) and editor of *Understanding Contemporary Asia Pacific* (Lynne Rienner 2007) and has written numerous articles and book chapters on ethnic politics in China. She served as Distinguished Fellow at Yunnan Nationalities Institute and Visiting Scholar at the Guangxi Ethnic Affairs Commission in China and as Senior Advisor on Minority Nationality Affairs for the US Congressional-Executive Commission on China. She is a National Committee on United States-China Relations Public Intellectuals Fellow.

Lyman A. Kellstedt is Professor of Political Science (Emeritus) at Wheaton College (IL). He received his B.S. from Bradley University and his Ph.D. from the University of Illinois. He is co-author or co-editor of the following books: *Rediscovering the Religious Factor in American Politics, Religion and the Culture Wars, The Bully Pulpit,* and *The Oxford Handbook of Religion and American Politics.* He is the author or co-author of articles in *American Journal of Political Science, Political Research Quarterly, Journal of the Scientific Study of Religion,* and *Social Forces,* as well as numerous book chapters. Upon retirement from Wheaton College, he taught for three years at Furman University and for one year at Texas A&M.

Akan Malici is Professor of Politics and International Affairs at Furman University. He is the author of *When Leaders Learn and When They Don't* (SUNY Press 2008) and *The Search for a Common European Foreign and Security Policy* (Palgrave Press 2008). Professor Malici is also the co-author of *U.S. Presidents and Foreign Policy Mistakes* (Stanford University Press 2011) and *Role Theory and Role Conflict in U.S.–Iran Relations: Enemies of Our Own Making* (Routledge 2016). He is the co-editor of *Re-Thinking Foreign Policy Analysis* (Routledge 2011) and *Political Science Research in Practice* (Routledge 2013, first edition). His

published articles appear in *Journal of Conflict Resolution, Journal of Peace Research, Political Psychology* and other journals. At Furman, he teaches classes in international politics and research methodology.

Melissa R. Michelson is Professor of Political Science at Menlo College. She received her B.A. from Columbia University and her Ph.D. from Yale University. She is co-author of *Mobilizing Inclusion: Redefining Citizenship through Get-Out-the-Vote Campaigns* (Yale 2012), *Living the Dream: New Immigration Policies and the Lives of Undocumented Latino Youth* (Paradigm 2014), *A Matter of Discretion: The Politics of Catholic Priests in the United States and Ireland* (Rowman & Littlefield 2017), and *Listen, We Need to Talk: How to Change Attitudes about LGBT Rights* (Oxford University Press 2017). Her work has also appeared in the *American Political Science Review, Journal of Politics, Political Behavior* and many other journals and edited collections.

Adrian D. Pantoja is Professor of Political Studies and Chicano Studies at Pitzer College. He has published extensively in the fields of Latino politics, immigration politics, and racial and ethnic politics. He is also a senior analyst with Latino Decisions, a public opinion polling firm that works with a number of political and advocacy groups seeking to understand the Latino electorate.

Elizabeth S. Smith is Professor and Chair of Politics and International Affairs at Furman University. She received her Ph.D. in American politics with a minor in political psychology from the University of Minnesota. She has taught research methods for over 20 years. She has published articles and chapters in *Journal of Political Science Education, Polity, Political Psychology, The Handbook for Teaching Social Issues, American Education Research Journal, Journal of Higher Education Outreach and Engagement* and in an edited volume entitled *Competition in Theory and Practice: A Multidisciplinary Approach* (Sense Publishers 2009). She is the co-editor of *Political Science Research in Practice* (Routledge 2013, first edition).

Brent J. Steele is the Francis D. Wormuth Presidential Chair and Professor of Political Science at the University of Utah. He is the author of *Alternative Accountabilities in Global Politics: The Scars of Violence* (Routledge 2013),

Defacing Power: The Aesthetics of Insecurity in Global Politics (University of Michigan Press 2010) and *Ontological Security in International Relations* (Routledge 2008). He is currently pursuing research projects on the topics of restraint, vicarious identity, micropolitics and further studies on ontological security. He is the co-editor of five books, has edited three journal special issues or symposia, and has published articles in a number of international studies journals, most recently in *Cooperation and Conflict, European Journal of International Relations* and *Millennium.* At the University of Utah, he teaches courses on US foreign policy, international security, international relations theory and international ethics.

Holly Teresi is currently a member of the Research and Market Intelligence team for Cox Automotive Inc. and has been a researcher for over a decade. Prior to joining Cox Automotive in 2014, Holly Teresi worked for multiple marketing research suppliers including Harris Interactive and CMI Research, focusing specifically on heavily regulated industries. Holly has also worked on nearly a dozen political campaigns, ranging from state senate races to presidential elections. She earned a B.A. in Public Communications from American University in Washington, DC and a Ph.D. in Political Science from Georgia State University. Her academic interests include political psychology, American voting behavior, technology and digital media, and experimental methods.

Kristina Thalhammer is Professor of Political Science and Director of Latin American Studies at St. Olaf College in Northfield, Minnesota. She is co-author of *Courageous Resistance: The Power of Ordinary People* (Palgrave Macmillan 2007) and has published articles in *Political Psychology, Perspectives on Politics* and *La Revista de Paz y Conflictos* among other publications. She earned her Ph.D. from the University of Minnesota and specializes in teaching comparative politics, with a focus on Latin America, human rights and political development.

C. Danielle Vinson is Professor of Political Science at Furman University. She is the author of *Congress and the Media* (Oxford University Press 2017) and *Local Media Coverage of Congress and Its Members* (Hampton Press 2003). She has published articles or chapters in *Political Communication, Harvard International Journal of Press/Politics, The Oxford Handbook of Religion and American Politics,*

and in the edited volumes, *New Directions in Media and Politics, Media Power, Media Politics* and *Blind Spot: When Journalists Don't Get Religion.*

Preface

Research methodology classes can be very difficult for teachers and students alike. Both of us have taught research methodology to undergraduate students for several years. We believe that we are well aware of the challenges and frustrations that come with teaching this class. We are also well aware of students' frustrations and their hesitations to take this class. An increasing number of political science departments in the United States and around the world, however, require their undergraduate students to take at least one research methodology class before they graduate. We are confident that this is a valuable requirement. Research skills are relevant, useful and applicable because students will need to utilize them at some point or even frequently in almost any career they may choose.

We saw the need for a new book mainly for two reasons: first, research methodology classes tend to be rather abstract. They are not topical and it is difficult to invoke students' interest in them. Students are naturally drawn to the field of political science because they can entertain interesting and engaging discussions about important and current topics such as contested elections, poverty, wars, *etc.* We agree with the students that political science should be first about the *Political* and second about the *Science.* All too often, however, we observe in our discipline that the *Science* takes prominence over the *Politics.* We also believe that typical methodology books, to undergraduate students, appear too divorced from actual politics and even from actual, real-world research. In this book we take a different approach to teaching methodology. We put the *Political* first and the *Science* second. We believe that any political science class should be guided by important political questions, problems or phenomena. The question then becomes *how* to study these issues.

This book is designed to expose students to the joy and rewards of research. In particular, we have asked excellent scholars in the field to use one of their

published scholarly works to show students how research is done. These scholars explain to students how they got interested in their particular research question, how they decided on what methodology would be most appropriate to answer that question, and what were the lessons they learned from engaging in this research. Students learn about the various methods used in political science by seeing it applied by these scholars. Unlike the typical methods textbook, students get a realistic picture of how the research process really works. They learn that it can be messy, frustrating, surprising but, most often, highly rewarding. We feel this approach is more engaging for students and does not set them up for the kind of unrealistic expectations (and, thus, frustrations) created by the standard textbook.

Our second reason for a new book is that we believe that typical research methodology books used at the undergraduate level have become too ambitious. We find that students get overwhelmed by the comprehensiveness, the extensiveness and complexity found in these books. If they remain undeterred nevertheless, we find too often that they get lost in the detail provided in typical research methodology books. For instructors, it is impossible to cover everything in these books, and for students it is difficult to discriminate and separate for each class meeting what they should focus on and what is of secondary relevance at this stage of their methodological training.

Even though in this book we give the *political* in political science more prominence, and even though our elaborations on the various research methodologies presented in this book are more succinct, we believe that students will learn better everything they need to learn about research methodology because they see how such methods are applied by excellent researchers. They are encouraged to engage actively with the real-world experience of the researcher as he/she explains the true story behind the scholarship. We hope that your experience will confirm this.

ACKNOWLEDGMENTS

Why Do We Need a Science of Politics?

Elizabeth S. Smith and Akan Malici

CONTENTS

In the late afternoon of election day, Tuesday, November 8, 2016, Nate Cohn, a political forecaster at *The New York Times*, had Hillary Clinton's chances of winning the presidential election at 85 percent. The Trump campaign, signaling what many thought was their sense that he was going to lose, had denied media

requests to film him and his team watching the returns. Early exit polls from that day had Clinton winning the electoral college with well above the requisite 270 votes. Conservative talk show host, Hugh Hewitt, a Trump voter himself, says about his prediction that Clinton would win, "I've never been this wrong." Clearly, he was not the only one.[1]

Hillary Clinton was an experienced candidate who had served eight years in the Senate. She was twice elected by the people of the state of New York and sat on the Senate Armed Services Committee. She was then appointed Secretary of State, a position she held for four years. She lost the 2016 presidential election to a businessman who, unlike all past winning presidential candidates, had never held either political office or served in the military. A candidate who refused to release his tax returns and who during the course of his campaign was revealed on a tape from a few years earlier to have been bragging about sexually assaulting women using the crudest of terms. During the election, Donald Trump questioned the heroism of a former Vietnam POW in his own party[2] and inspired a group of prominent Republican foreign policy experts to issue a letter calling him "fundamentally dishonest" and "utterly unfitted to the office."[3]

Given the stark differences in experience between the two candidates and the unprecedented features of Donald Trump's candidacy, most had *assumed* that though Hillary Clinton was a flawed candidate she would win the election. The questions on many minds in the days and months following the election was: How did Donald Trump win? And, why did the pollsters and pundits get it so wrong?

The high-school geometry teacher of one of us (Elizabeth Smith) once printed the word ASSUME on the board, and to our teenage amusement pointed out that when you ASSUME you are making an "ass" of "u" and "me." Quite cleverly, she illustrated why in fact we need science. Our assumptions, our hunches, our general gut level feelings about the "truths" of the world are often proven wrong once we engage in careful, systematic and scientific analysis.

Science: A Way to Know

To know, according to Webster's, is "to have a clear perception or understanding of; to be sure or well informed about." There are various types of knowledge. Among them are religious knowledge, instinctual knowledge, common sense, scientific knowledge and so on. Although our book is about scientific knowledge we do want to point out that scientific knowledge is not necessarily better than the other types of knowledge. A mother's intuition about what to do with her unsettled child may trump the conclusions of scientific studies many times, for example. We value scientific knowledge, but we believe it is important to retain an appreciation also for other types of knowledge.

Science is a distinct type of knowledge and it stands apart from any other type of knowledge. The noun "science" originates from the Latin verb "scire," and translates into "knowing objectively," more specifically it is a way of knowing that is systematic, **replicable, cumulative** and **falsifiable**. In other words, scientific knowledge is based on careful and comprehensive observation of the data (systematic), collected and analyzed in such a way that others can reproduce the analysis (replicable). Scientific knowledge often evolves over time as multiple methodologies are being used and new data is examined (**cumulative**). Finally, scientific knowledge must be open to questioning and the possibility of being disproven by new data (falsifiable). The election of Donald Trump is one we can begin to understand more completely by looking at careful scientific analysis.

While there are many features of the 2016 presidential election scholars are still studying, a body of scholarship prior to the election provides important insight into Donald Trump's win. Headlines explaining Trump's victory immediately after the election suggest that Trump won because of the economy – displaced workers in the new economy found Trump's message of protectionism and bringing back jobs appealing, so the argument went. Journalists wrote stories like the one by Tami Luhby on CNN Money suggesting that Trump won because he was seen by voters as more likely to improve the economy. Luhby reports that 49 percent of voters said Trump "would better handle the economy," while only 46 percent said Clinton would.[4]

However, scholarly research suggests that Trump's election can be better explained by examining the cultural stressors pushing voters toward populist appeals. Inglehart and Norris, for example, use multiple methodologies, and they provide compelling evidence that in both the US and Europe economic concerns

of voters are trumped (pun intended) by fears raised by the perceived erosion of stature and identity, especially among traditional white voters, by increasing immigration, perceived loosening of traditional moral values and policies promoting gender and marriage equality.[5] Similarly, political scientist Philip Klinkner, who uses the 2016 *American National Election Study (ANES)*, a representative study conducted since 1948 by political scientists trained in survey research methods who are interested in understanding voting behavior, provides further insight into Trump's win. In particular, examining closely the *ANES* pilot survey of voters before the election, he shows that racial resentment, the feeling that blacks and other minorities were given unfair advantages, better predicts support for Trump than do economic concerns.[6]

A careful reading of the journalistic report cited by Luhby of CNN Money would raise red flags in a student such as yourself who is (soon to be) well trained in research methodology. Political scientists and other statisticians KNOW that the 49 percent to 46 percent difference in reported voter support for Trump's versus Clinton's ability to handle the economy is well within the range of a typical **margin of error**, the difference between the sample and the **population** value when only a sample of people are surveyed. A margin of error is applied to the spread of the percentages reported, meaning that a +/−3 percent margin of error applied to these percentages could mean that the actual poll results could be a virtual tie (49% to 49%) or as large a difference as 52 percent (49% + 3%) to 43 percent (46% − 3%), or could be flipped (46% for Trump and 49% for Clinton). All of these statistics are very different in terms of potential electoral consequences, but all within the margin of error. In this particular case, the report presented on CNN Money is even more difficult to interpret accurately as, unlike in a scholarly, scientific study, no information is provided to the reader on how the data was collected (it was simply reported that it was a CNN exit poll), what the sample size and margin of error was, and whether it included a true *random sample* of the *population*, where no member of the population has any more chance than another of being included. Thus, this study is non-scientific in nature in large part because it is neither *systematic, verifiable* nor *replicable*, at least in regard to the form in which we the readers see it online.

But, we are still left with the question: Have political scientists, who point to the cultural influences on Trump voters, discovered the absolute "truth" about

why Donald Trump won the 2016 election? Scientists would caution that scientific findings do not indicate absolute certainty if that is what we mean by "truths." Instead, scientists tell us that any scientific knowledge is in fact tentative. Scientists have ensured themselves job security in the fact that it is an enterprise based on continual investigation of new data in new ways. As the evidence accumulates we certainly feel more confident about our conclusions, but any good scientist knows that their conclusions could one day be disproven by the accumulation of contrary scientific evidence. That is why most of us find science such fun! We are constantly exploring our surroundings, making careful observations and testing and re-testing our hunches in a search for knowledge and understanding of the complex and interesting world in which we live.

The Scientific Study of Politics: The Birth of the Discipline of Political Science

It is easy to think of false truths in the natural world that have been disproven by systematic, cumulative, scientific evidence: the false truth that the world was flat; the false truth that maggots were spontaneously generated in decaying meat; the false truth that the sun revolved around the earth; the false truth that a heavy object travels faster through space than a lighter one (Elizabeth still finds that one hard to let go of). As with the misunderstandings of the early explanations of support for Trump, the world of politics has been plagued by false truths.

False truths exist everywhere, including the political world, and for this reason it was suggested that a science of politics be created. In fact, in 1903, the **American Political Science Association** was formally created in an attempt to create an objective, systematic scientific study of claims made about the political world. The address to the association's members by the first president of the American Political Science Association, Frank J. Goodnow, emphasized that "scientific knowledge provided a check on the tendencies of political theory 'to soar in the empyrean realms of speculation'."[7]

Prior to the creation of this association, the study of politics was often **normative**, meaning it was concerned more often with theorizing about how

things *ought* to be rather than on understanding how things actually *are.* Students of politics began to understand that one important way to credibly make claims about how things ought to be would be to have a better understanding of how they actually are (a **positivist** approach). The discipline gradually evolved over time. At first, there was what could be labeled **traditional political science**. It was concerned mainly with documenting the institutions of government and the laws, rules and norms of the politics. It was idiographic and descriptive rather than nomothetic and analytical: it was largely limited to a description of individual cases or events rather than the discovery of general laws pertaining to many cases or events. Yet, the ambition to discover generalities and cause–effect relationships grew over time, and along with it grew more advanced and precise scientific methods and tools (like survey research, statistical analysis aided by computers, etc.). Beginning around the 1950s, these advances came to be known as the **behavioral revolution** in political science. Two of the main traits of this revolution were the quantification and the formalization of the study of politics toward a more precise, empirical and explanatory science.

The behavioral revolution helped our discipline become more scientific, yet it also led to **postbehavioral** criticism. This criticism asserted that political and social conduct cannot be quantified and analyzed through mathematical **models**. There was also the criticism that political scientists became more concerned with sophisticated methodologies than the substantive issues that they should care about. Today most political science studies are marked by the behavioral revolution, some more so and some a little less.

Political science, as a field, is divided into four major subfields: **American politics**, **comparative politics**, **international relations** and **political theory**. Unlike the first three subfields, political theory does not depend on **empiricism** (the observation of the data) but instead relies on logical, discursive consideration of ideas without systematically examining their validity. Because the process of political theory research is fundamentally different from the empirical subfields of American politics, comparative politics and international relations, we will not discuss it in this textbook. However, the fact that we will not discuss political theory in this textbook should not be taken as an indication of its lack of importance. The thoughts and theories set forth by political theorists have been extremely important to the study of politics and to the creation of our political

institutions. In addition, the ideas of political theorists are often the basis for the ideas being tested in systematic, scientific ways by the other subfields which do rely on empirical data.

Let's consider a few of the many examples of the ways scientific knowledge accumulated by political scientists has disproven our assumptions regarding what is true in politics in each of these three subfields: American politics, comparative politics and international relations.

The Importance of Science: Debunking False Truth in American Politics

The subfield of **American politics** is concerned with the study of political institutions and behavior in the American political system. Scientific study of American politics has helped debunk some common assumptions made by both the people and the press regarding how things really work in the American political system. For example, consider the problem of women's underrepresentation in elected office in the United States. Despite the fact that women make up over 50 percent of the US population, in our entire nation's history, of the 12,000 or so members of Congress who have ever served, only about 2 percent of them have been female.[8] The United States ranks 100th in the world for representation of women in the national legislature.[9] It has long been an assumption that one significant reason that women are underrepresented in politics is because voters are biased against females and unwilling to vote for them. Many used this argument to explain why Hillary Clinton lost to Donald Trump, including Hillary Clinton herself after the election.[10] Public opinion polling also tells us that this is what most voters think. As Lynne Ford, a political scientist who studies women and politics points out, one poll found that "two-thirds of voters believed that women have a tougher time winning elections than men do."[11] Another study found that potential female candidates also believe that there is a sex bias by voters against female political candidates.[12] A significant number of scientific studies by political scientists, however, have disproven this assumption. In fact, all other things being equal (like professional background,

financial resources and political context), female candidates fare as well, and in certain situations, fare even better than their male counterparts.[13] A recent careful examination using multiple methodologies (including content analysis, a method discussed in Chapter 10 of this book) of congressional elections in 2010 and 2014 shows that women are not subject to biased media treatment nor punished for being women by voters.[14]

False assumptions can have important, real-life consequences. In this particular case, scholarly research shows that these false assumptions in fact perpetuate the problem of women's underrepresentation, as qualified women decide not to run, believing that voters and the media will treat them in an unfair and biased way. Political scientists interested in women's underrepresentation have a responsibility to share the truth as we know false truths affect our behavior and decisions, most especially in this case, the choices made by qualified, potential female candidates.

The Importance of Science: Debunking False Truth in Comparative Politics

Let's consider some examples of the importance of science in addressing false assumptions in the subfield of political science known as **comparative politics**. While the subfield of **American politics** is concerned with studying the institutions and behavior of the American political system, comparative politics is the study of the differences and similarities in institutions and processes of politics across political entities, often involving comparisons of different nation-states. Political scientists in this subfield use what is called the **comparative method** to understand and explain the political world (see the discussion in Chapter 3). The comparative method may involve one of two approaches: (1) comparing nations (or any other political entities of interest) that are similar in all respects except for one, to see the cause/effect of that one difference (called the **method of difference**); or (2) comparing nations (or any other political entities of interest) that are different in all ways except for one to understand the cause/effect of that difference (called the **method of similarity**).

One question that has been of interest to political scientists since really the dawning of the field is the question: "Why do some [new] democratic governments succeed and others fail?"[15] One common assumption or theory held by scholars was that economic prosperity was an essential (if not the most important) component to democratic success. Many scholars have asserted that economic development and modernization are fundamental to the creation and maintenance of a successful democracy.[16] In a fascinating comparative analysis, however, Robert Putnam finds very interesting evidence showing that economic prosperity and development may not be the determining factor it was once thought to be in shaping effective democracies. Putnam uses the method of similarity, comparing 20 different regional governments in Italy which though very different on a variety of variables, especially in the Northern versus the Southern regions, were all similar in one important respect – they were all created in a top-down way with similar structures and rules by the central government in 1948 under a new Italian constitution. Putnam and his colleagues, using numerous scientific methodologies, including personal interviews with political elites (see Chapter 5 for a discussion of this method), multiple nationwide surveys (see Chapter 8), analysis of statistical data (see Chapter 7), and even a field experiment (see Chapter 11), find that contrary to the popular belief that economic development predicts democratic success, it is instead a long history of civicism in a region, including civic engagement and social capital, defined as "trust, norms and networks," that is most important in predicting democratic institutional success.[17]

Putnam's work received a lot of attention and praise precisely because its innovative, systematic and scientific approach to an issue long felt settled helped upend traditional scholarly thinking on the question of what makes for effective democratic institutions. Thus, scientific knowledge not only provides clarity for false popular and media assumptions but also for false scholarly assumptions.

The Importance of Science: Debunking False Truth in International Relations

Scientific knowledge about the political world is important, not only to overcome faulty assumptions made by the public, the media and scholars, but also to overcome faulty assumptions made by politicians. Let's take an example in the subfield of political science known as **international relations** to see how this has worked. International relations is the study of the relationships between or among the various states and other political entities of the world, such as governmental or non-governmental international organizations. Most, but not all, international relations theory is what scholars call **positivist** – in that it relies on the scientific method and empirical data to help explain and understand the relationships among political entities. Let's consider one assumption made by politicians regarding how nations interact with each other which international relations scholars have since proven wrong by using scientific analysis, or a positivist approach.

When thinking about the security of one's nation, politicians have often concluded that building up arms, and establishing military alliances and defense systems should make a nation safer. On the face of it, this belief appears logical as more weapons and a stronger military with more alliances provide protection against aggression by one's enemies. As a defensive security strategy, the thinking goes, a buildup of arms and the creation of alliances should not worry other nations unless they already have the intention of acting aggressively toward us. Consider this exchange by Senator Tom Connally and Secretary of State Dean Acheson regarding the ratification of NATO, which was an attempt to bolster the defenses of cooperating nations against aggression, particularly by the Soviet Union. This exchange shows this kind of logic regarding how military buildup will be interpreted by other nations being used by politicians:

> **Secretary Atcheson**: [The treaty] is aimed solely at armed aggression.
> **Senator Connally**: In other words, unless a nation … contemplates, meditates, or makes plans looking toward aggression or armed attack on another nation, it has no cause to fear this treaty.[18]

These politicians believed that only a nefarious nation would be concerned with another country's purely defensive actions. As they go on to discuss in this exchange, just as only a person likely to commit crimes needs to worry about stricter crime laws, only a nation with evil intent needs to worry about another

nation's buildup of military arms and defenses.

A leading scholar in the field of international relations by the name of Robert Jervis, however, uses a scientific methodology known as **formal modeling** to disprove this theory. Formal modeling is a method whereby a mathematical and logical model, coupled with examination of real-world politics, is used to examine relationships, and predict and test interactions among entities. Using this technique, Jervis develops what is known as the theory of the security dilemma. The theory starts with the assumption of anarchy. Anarchy, when used by international relations theorists, does not mean "chaos." Rather it refers to the absence of hierarchy in the international system: above states, there is no authority such as an overarching governing body (a world government) that can enforce how nation-states interact with one another. A domestic system, by contrast, is organized hierarchically: there is an authority (the court system) that is able to enforce how citizens of this state interact with each other.

Jervis shows that, in fact, in an anarchic international system one nation's defensive move to build up arms and form alliances will likely be interpreted by other nations not as defensive but as aggressive, thus leading that nation to build up its military defenses and arms, eventually leading to an even more precarious situation for all. In other words, as one country attempts to be more secure, they are actually instigating a series of responses from others which in the long run make them less secure than they were to begin with. Because nation-states cannot assume that the intentions of a nation are merely defensive, they must react with more weapons and defenses themselves. Nation-states become less willing to work with one another as the fear and mistrust grows in response to the buildup. Thus, the threat of aggressive interactions between the two sides increases rather than decreases as intended. Misunderstanding of the security dilemma, Jervis states, can have important political consequences as politicians fail to consider other means (such as diplomacy) which might actually be more likely to result in enhanced security for a nation than would a buildup of arms and defenses. Scholarly research on the security dilemma has been used to inform politicians in their considerations of the best security strategies.

The Naïve Scientist and Beyond

You may have heard the term "naïve scientist" before. Political scientists use it sometimes, but it originates from our cognate discipline psychology and there is a fascinating insight behind it if you think about it. Before any of us receive any methodological and scientific training we are all naïve scientists. As laypeople we are "scientists" because we perform exactly the same tasks as trained scientists do: we collect data (through our observations and experiences), we observe **correlations** between different observations, we make **generalizations** from these observations, and we also formulate causal hypotheses and theories to predict events. Finally, we test these theories against new observations. Yet we are naïve, because as human beings, we are biased in many ways and we are not following a rigorous and systematic logic (method) when going through these steps. The very simple goal of this book is to help us become less naïve and more methodological – as everyday citizens and as political scientists.

Being methodological in answering our questions is important. Consider the following questions: Does democratization lead to peaceful foreign relations? Do diverse cultures necessarily clash with each other? Does religion impact politics? If so, in what ways? Do "values" have an impact on politics? Do women have equal chances in elections? Does the socioeconomic class we're born into impact or even determine our later prospects in life? These are all interesting and indeed very important research questions. The popular media and journalistic accounts of domestic and international accounts have led to many "false truths" in answering these and other important questions. However, both citizens and political scientists have an interest in acquiring "truth statements" about such questions. Political scientists especially have a responsibility to provide such knowledge, as they are to communicate their results not only to a community of scholars but also to all people. The competence in research methodology justifies the making of statements that help constitute general or particular bodies of knowledge justifiably described as "truthful" or "scientific."

In this textbook, you will be introduced to the scientific study of politics and how it is conducted through a conversation with scholars in the field. These scholars will expose you to the various methodologies by discussing precisely how they themselves went about engaging in the research process. Each chapter's author(s) will use one of their own published studies to show you an example of

one of the variety of methodologies available to us to answer important political questions. These methodologies include the comparative case study method (Chapter 3), field research (Chapter 4), interviews (Chapter 5), critical and interpretive analysis (Chapter 6), statistical research (Chapter 7), survey research (Chapter 8), public policy and program evaluation (Chapter 9), content analysis (Chapter 10), and field experiments (Chapter 11). These methodologies range from the more qualitative (non-numerical, usually done with a smaller sample of data and often more hypothesis-generating than hypothesis-testing methods such as comparative **case studies** (see Chapter 3) to the more quantitative (numerical and usually involving a larger sample of data such as **statistical analysis** (see Chapter 7)).

You will hear firsthand from the political scientists who conducted this research. You will discover why they got interested in their particular research question, what challenges they faced in answering it, what methodologies they found were most appropriate to answer their question, and what "truths" they were able to discover. You will see how being a scientist is much like being a detective. Solving a problem or political mystery requires information-gathering and analysis in a systematic and thorough way. You will see that scholars are motivated by an unanswered question or a tension they feel regarding what is assumed to be true and what they have observed to be true. And, you will hopefully learn not just about the methodologies available to you but also about the fun you can have applying these methodologies to the search for truth about politics.

Exercises and Discussion Questions:

1. Do you think that *religious knowledge* and ***scientific knowledge*** are reconcilable?
2. Do you think the behavioral revolution in political science is to be welcomed? What are the benefits it has produced? And what are some pitfalls it has created?
3. Identify a situation in which you were a "naïve scientist." Describe

yourself in that situation. How could you be less "naïve" in this situation?

4. Identify a claim about the political world made by the media or political pundits (for example, Congress is more partisan than ever; or, Islam is incompatible with democracy). Then, find scholarly research in political science that addresses that claim. Is the claim true according to this research? Does the scientific data support the media claim? What methodology does the scholarly research use to evaluate this claim scientifically?

Recommended Resources:

The American Political Science Association (www.apsanet.org): This website will provide you with information about the primary professional association for political scientists. You can find out about the history of the field, can access links to journal articles and read about the important activities in which political scientists are engaged.

The New York Times (www.nytimes.com): Check out one of the most comprehensive media outlets for coverage of both domestic and international politics.

Politico (www.politico.com): For some of the most current discussion of American politics by journalists, check out this political website.

Gallup (gallup.com): A reputable polling organization using scientific sampling methods to assess public opinion on the most pressing political issues of the day.

Survey Documentation and Analysis (www.sda.berkeley.edu): This website provides you with access to some of the best academic surveys in political science including the General Social Survey and the National Election Study. The website allows you to select survey questions of interest to you and do basic statistical analyses with them.

Notes

1 Stelter 2017.

2 Schreckinger 2015.

3 War on the Rocks Staff 2016.

4 Luhby 2016.

5 Inglehart and Norris 2016.

6 Klinkner 2016.

7 Gunnell 2006: 483.

8 Center for American Women and Politics 2010.

9 Catalyst 2017.

10 Robbins 2017.

11 Ford 2011: 140.

12 Lawless and Fox 2010.

13 Seltzer, Newman and Leighton 1997.

14 Hayes and Lawless 2016.

15 Putnam 1993: 3.

16 See, for example, Dahl 1971; Lipset 1960.

17 Putnam 1993: 167.

18 Cited in Jervis 1978: 181.

References

Catalyst. 2017. "Women in Government." www.catalyst.org/knowledge/women-government.

Center for the American Woman and Politics. www.cawp.rutgers.edu/.

Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition.* New Haven: Yale

University Press Ford, Lynne. 2011. *Women and Politics: The Pursuit of Equality.* Boston: Wadsworth.

Gunnell, John G. 2006. "The Founding of the American Political Science Association:

Discipline, Profession, Political Theory and Politics." *American Political Science Review* 100(4): 479–486.

Hayes, Danny, and Jennifer L. Lawless. 2016. *Women on the Run: Gender, Media and Political Campaigns in a Polarized Era.* New York: Cambridge University Press.

Inglehart, Ronald F., and Pippa Norris. 2016 (August). "Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash." Faculty Research Working Paper Series, Harvard Kennedy School of Government.

Jervis, Robert. 1978. "Cooperation Under Security Dilemma." *World Politics* 30(2): 176–214.

Klinkner, Philip. 2016 (June 2). "The Easiest Way to Guess if Someone Supports Trump? Ask if Obama is a Muslim." https://www.vox.com/2016/6/2/11833548/donald-trump-support-race-religion-economy.

Lawless, Jennifer L., and Richard L. Fox. 2010. *It Still Takes A Candidate: Why Women Don't Run for Office.* Cambridge: Cambridge University Press.

Lipset, Seymour. 1960. *Political Man: The Social Bases of Politics.* Garden City, NY: Doubleday.

Luhby, Tami. 2016 (November 9). "Voters Trust Trump to Fix the Economy." http://money.cnn.com/2016/11/09/news/economy/trump-economy-voters/.

Putnam, Robert. 1993. *Making Democracy Work.* Princeton, NJ: Princeton University Press.

Robbins, Mel. 2017 (May 3). "Hillary Clinton Lost Because of Sexism." www.cnn.com/2017/05/03/opinions/hillary-clinton-interview-sexism-robbins/.

Schreckinger, Ben. 2015 (July 18). "Trump Attacks McCain: 'I Like People Who Weren't Captured.'" www.politico.com/story/2015/07/trump-attacks-mccain-i-like-people-who-werent-captured-120317.

Seltzer, Richard A., Jody Newman, and Melissa Vorhees Leighton. 1997. *Sex as a Politcal Variable: Women as Candidates and Voters in U.S. Elections.* Bouder,

CO: Lynne Reiner Publishers.

Stelter, Brian. 2017. "In Their Own Words: The Story of Covering Election Night 2016." http://money.cnn.com/2017/01/05/media/election-night-news-coverage-oral-history/.

War on the Rocks Staff. 2016 (March 2). "Open Letter on Donald Trump from GOP National Security Leaders." https://warontherocks.com/2016/03/open-letter-on-donald-trump-from-gop-national-security-leaders/.

How Do We Get a Science of Politics?

Akan Malici and Elizabeth S. Smith

CONTENTS

Not too long ago the well-known Harvard University international relations scholar Stephen Walt wrote the following on the popular website foreignpolicy.com: "Every time the U.S. touches the Middle East, it makes things worse. It's time to walk away and not look back."[1] Indeed, the balance sheet of US engagement in the Middle East does not look good. Walt argues that our foreign policymaking elite are one reason the US continues to put itself in trouble in the Middle East (and elsewhere). Too many members of this elite have a naïve belief in some liberal international relations theories. The word "liberal" in the preceding sentence is not to be confused with the conventional way in which it is used – an ideology associated with political Democrats. Instead, it refers to a particular strand of international relations theories and the policy prescriptions

following from these theories, which have at times been supported by Republicans and Democrats alike.

One of these prominent liberal international relations theories is the Democratic Peace Theory. What does this theory say? It says, in short, that democracies are peaceful with each other; they don't go to war with each other. Therefore, it seems to be in the interest of the US that authoritarian states become democracies, especially those states that can affect US interests. For American leaders then, this becomes the task of nation-building and this is what they have tried to do in recent years in Afghanistan and Iraq, for example.

The Democratic Peace Theory is a theory developed by political scientists and employed by policymakers. Many would say this is how it should be – that there is a bridge between academics and policymakers and that both work together toward good policy. Whether good policy is indeed the outcome is an open question and we'll return to it again in the conclusion of this chapter. The main thrust of this chapter concerns the question of how we get a science of politics. The answer is by adopting a language of science and by proceeding systematically and scientifically in one's inquiry of politics. In this chapter we begin by providing a basic language of science as we trace step by step the journey of the Democratic Peace Theory from its inception to its influence in the highest offices of policymaking in Washington, DC. In doing so, we shall highlight those terms that will be important for you while reading and studying this book.

Getting Curious

The origins of the Democratic Peace Theory are very interesting. They date back more than 200 years. It all started with the observations of the German philosopher Immanuel Kant in his hometown of Konigsberg, which was then the capital of Prussia. Kant was very concerned about all the conflict he observed all around the world. Naturally for a philosopher, he thought a lot about this issue and he wrote down his thoughts in a little book that was first published in 1795. The title of the book was *Perpetual Peace.*

The phenomenon that Kant set out to study was world peace and he asked a fundamental **research question**. A research question is a statement that identifies the phenomenon we want to study. It generally is motivated by our curiosity about something that we consider important but that has not been asked, addressed and answered yet – at least not satisfactorily. Kant's research question was: How could the world attain peace? Even more ambitious, as the title of his book suggests, he asked how the world could attain perpetual peace, meaning peace that would last into eternity. Kant's question was indeed a big question. But his answer is relatively simple. Here it is:

[The] republican constitution … provided for this desirable result, namely perpetual peace, and the reason for this is as follows: If (as must inevitably be the case, given this form of constitution) the consent of the citizenry is required in order to determine whether or not there will be a war, it is natural that they consider all its calamities before committing themselves to so risky a game. (Among these are doing the fighting themselves, paying the costs of war from their own resources, having to repair at great sacrifice the war's devastation, and finally, the ultimate evil that would make peace itself better, never being able – because of new and constant wars – to expunge the burden of debt). By contrast, under a nonrepublican constitution, whose subjects are not citizens, the easiest thing in the world is to declare war. Here the ruler is not a fellow citizen, but the nation's owner, and war does not affect his table, his hunt, his pleasure, his court festivals, and so on. Thus, he can decide to go to war for the most meaningless of reasons, as if it were a kind of pleasure party, and he can blithely leave its justification (which decency requires) to his diplomatic corps, who are always prepared for such exercises.[2]

In short, Kant is saying that if a country is a republic, then it will be very hesitant to go to war. Notice that *republic* is a Latin term (*res-publica*) and translates literally into "the affairs of the public." Semantically it is very similar to the term *democracy* which has Greek roots (*demos kratein*) and translates literally into "the rule of the people." Thus, in more familiar language, we may say that Kant advanced the **proposition** that democracies are unlikely to go to

war. A proposition is an expression of a judgment or a declaration about the relationship of at least two **concepts**. The concepts in this proposition are *democracy* and *war*. **Concepts** are the words we choose to describe the phenomena we are interested in.

Concepts must be **conceptualized**, which simply means that they must be defined. In other words: What is a war? What is a democracy? Let's focus on the latter concept – democracy. Some adopt a narrow definition, emphasizing the competition of at least two parties and free elections. Others add on and say that for a country to qualify as a democracy it must also have and guarantee fundamental freedoms such as the freedom of speech, the freedom of religion and a free media. It will be obvious to you that if researchers adopt differing **conceptualizations** for their concepts, then they are likely to use different indicators (or precise measures) of their concepts. The process of denoting the indicators of a concept is known as **operationalization** of that concept. The strength of the relationship between a concept and its indicator is known as the **epistemic correlation**. So, for example, an IQ test is often used as an indicator of the concept intelligence, although some would make the argument that it is a poor indicator of that concept (they would be arguing that the epistemic correlation between the indicator and the concept is weak). If operationalizations differ among scholars, then the research findings are likely to be different. As a result, the comparability of their research is jettisoned – yet, comparability should be a central goal of any research community.

An important next step in the research process is to ask: Is the proposition true? So, in this example, is the proposition that democracies are less likely to go to war true? By definition, **propositions** must be either true or false. Immanuel Kant was a philosopher and his proposition appeared plausible, but at his time there was no **empirical evidence** to support it. Empirical evidence is proof that a proposition is true. In the social sciences we often like to have **data** as proof. Data is nothing more than systematically collected and objective observations about the phenomenon we are studying. It can come in two forms: quantitative or qualitative (numerical or textual). If we have data confirming a proposition, then we gain more confidence in it. The proposition then also becomes more relevant for it may carry implications on our actions and doings. For example, if Kant's proposition is indeed true, then one important implication would be that the

establishment of democratic governments should be actively promoted as a way to minimize conflict and promote peace. In fact, those among you who have taken an interest in international politics will know that, as we alluded to in the introductory paragraphs of this chapter, various US administrations, Republican and Democratic ones alike, have made such implications a driver of their foreign policy.

Working Toward an Answer

It was not until about 150 years later, in the 1950s and 1960s, that political scientists embarked upon delivering empirical evidence for Immanuel Kant's proposition. These researchers proceeded step by step and very systematically. Let us go through and explain each step. The researchers began by articulating two specific research questions that are derived from Kant's proposition, but they are a little more specific. Their **research questions** were twofold: 1) Are democracies more peaceful in general? 2) Are democracies more peaceful only toward other democracies?

Here we reach a good opportunity to elaborate a little more on what a good or appropriate **research question** should look like. First, a research question should be a **positive question**. This does not mean that it should be a 'good' question. It means that the question should be about phenomena in the existing world. The opposite of a positive question is a **normative question**. Normative questions are about how things ought to be in an imagined world. An example of a normative question would be: How should a country be governed? A good way to think about positive questions versus normative questions is that the first address a 'What *is*?' question, whereas the latter address a 'What *ought* (to be)?' question. This does not mean that positive questions cannot have normative implications. In fact, many would say they should. We will return to this issue at the end of this chapter and also at the end of this book.

Second, **research questions** should not be about single factual issues, such as the question, "How high or low is the voter turnout in national elections in country X?" There is not much to be researched here – the answer can be looked

up very quickly. Perhaps more importantly though, there is nothing puzzling behind this question. Instead, research questions should be about relationships or associations between at least two issues (concepts). An example would be, "Do negative campaign ads lead to a depression in voter turnout?" The two issues here are a) the nature of campaign ads and b) the voter turnout. One way of thinking about this second point is that research questions should be less "What?" questions, but rather "Why?" questions. In the present example, we are seeking an answer to *why* there may be low participation in national elections.

The third and fourth point for a good research question can quickly be stated together. The third point is that research questions should be concrete; that is, their scope should be clearly defined. Our example about the nature of campaign ads could perhaps be revised to read "Do negative campaign ads *on television* lead to a depression in voter turnout?" This formulation makes it clearer what the researchers are about to investigate – the question is concerned specifically with television ads. Data should be available towards answering the posed research question. Indeed, voter participation in national elections is often available online and negative campaign ads on television can also easily be quantified (counted) by a research team. Fourth, research questions should be subjected to the "So what?" test. The "So what?" test ensures that the answers we will gain from engaging in these questions should be relevant and have implications that go beyond the bounds of a particular research project.

After stating their research questions, the researchers often formulate **hypotheses** to guide their research. Hypotheses are very similar to propositions and oftentimes a hypothesis can be derived from either a proposition or a research question. Hypotheses are expected, but not yet confirmed relationships between two or more variables. A good hypothesis should meet several criteria. First, like a research question, it should be a **positive statement**. That is, it should be an informed guess about an expected relationship in the existing world. Second, a good hypothesis should be a general statement. This means it should not just address a single event, such as the cause(s) of a *particular* election with a low voter turnout, but it should address the cause(s) for low voter turnout, *generally*. Third, it should be a plausible statement. This means there should be some good reason why we believe that a proposed relationship is true (this is why we refer to a hypothesis as an 'educated guess'). Fourth, again like a

research question, it should be a testable statement. This means data should be available so we can assess whether the hypothesis can be confirmed or must be disconfirmed.

All of the steps we have discussed so far, and also many of those that follow, require the researcher to do a **literature review**. A literature review is best explained by laying out its purposes. The two fundamental ones are: a) to establish the state of knowledge in the issue area that the researcher wants to explore, and b) to identify the research space where s/he intends to join the scholarly discussion and contribute to it. What exactly does this mean? In [Chapter 1](), we said that science progresses cumulatively. In order to know where one can make a contribution, one needs to know what aspects of an issue have been investigated already and what has not been investigated yet. Toward this end, the researcher will read as many books, articles and other forms of publications on the issue as s/he can. Doing so will allow the researcher to identify the questions that have not been answered yet, or at least not satisfactorily. It is these answers that will make an original contribution to the field. This is what good researchers aspire to do.

A literature review has yet further purposes. In the process of doing a literature review, the researcher will be able to identify the important **concepts** in the subject area as well as possible relationships among them. The researcher will learn how previous scholars have **conceptualized** (defined) and **operationalized** (measured) these concepts and what data sources have been used. As the researcher identifies relationships among concepts, s/he will also be able to identify new and original hypotheses that can be tested. All of this allows the researcher to proceed contextually and in an informed way in his/her scholarship.

When you engage in a research project, you will also have to engage in a literature review. Once you have identified your research question and the scope of your investigation, you will need to identify books and articles pertaining to your topic. These should be written by credible and authoritative authors and published in recognized and authoritative outlets. Identifying this literature is not always easy. How does one know what is a good publication and what is not? There are various markers that will help with answering this question. For articles, for example, one marker of quality control is whether they are **peer-**

reviewed. If an article is peer-reviewed this means that it has gone through a rigorous process of review undertaken by other expert scholars. These reviewers will check the article for its theoretical premises, the appropriateness of the used methodologies, the **validity** and **reliability** of the data, the firmness of the conclusions and also the overall appearance of the article. Once this is done and the article is published, the reader can generally be confident that the publication is of high quality. Another marker of quality control is the credentials of the author. Does s/he have relevant schooling and a degree pertaining to the subject? Has s/he written before on the subject matter? If s/he has not, that should not disqualify his/her research from being important to you. But it will be good for you to know who the established figures are in any given field versus who the newcomers are.

There are other markers for quality control, and you will learn many of them as you progress through your studies. Honestly, some of them are learned simply through experience, and the more experienced a researcher is, the better versed s/he can be in the literature review. You are now laying the foundation for this experience. For your next research project you should definitely work with a reference librarian on your campus's library.

Now let's return to the Democratic Peace Theory. After having stated their research questions, the Democratic Peace researchers formulated hypotheses. These corresponded to the two research questions on page 17. The hypotheses were:

- *If a regime is a democracy, it will engage in peaceful crisis behavior.* This hypothesis was called the monadic version. It is called monadic because it asserts that a democracy will engage in peaceful crisis behavior (not resort to force) regardless of whether the other regime is a democracy or not.
- *If two democratic regimes encounter each other in a crisis, they will resolve it peacefully.* This was the dyadic version. It is called dyadic because it asserts that a settlement will be the outcome only if both contending sides are democracies. In other words, it takes two democracies to make peace.

The **unit of analysis** in these hypotheses is countries. A unit of analysis is

simply the object or the entity under study. In our example, our study is concerned with states and their behavior (towards other states) in crises situations. In other studies, such as the voting behavior of individuals in national elections, the unit of analysis would be, as just suggested, individuals. If the study would examine the voting behavior of religious denominations, the unit of analysis would be (religious) groups.

By formulating hypotheses, the researchers express the purported cause and effect in scientific terms. Across all sciences the presumed cause is referred to as the **independent variable** (often referred to as IV) and the presumed effect is referred to as **dependent variable** (often referred to as DV). The independent variable was type of regime and the dependent variable was crisis behavior for the monadic hypothesis and crisis outcome for the dyadic hypothesis. We refer to independent and dependent variables as variables because they can vary, that is because they can take on different variations or values. In fact, any variable must be able to assume at least two distinct values. Otherwise, by definition, it is not a variable. The independent variable can take on the values democratic or non-democratic. The dependent variable can take on the values peaceful or conflictual behavior (in the monadic version) and settlement or war (in the dyadic version).

A third type of variable is an **antecedent variable**. An antecedent variable occurs in time prior to an independent variable and may act as a catalyst for the independent variable. What could be an antecedent variable in our case? Or, let's ask more specifically: What could be an antecedent variable to democracy? A careful review of the literature would show us that scholars of democratization have identified a host of factors. Among them are social and political pluralism, high levels of literacy and education, traditions of toleration and compromise, economic development and social modernization. A market economy is considered one of the most prominent factors towards generating a democratic regime in a country. So we could say that a market economy, or more specifically a liberal and open economy, is an **antecedent variable** (or a precondition) to democracy.

Researchers embarked upon **quantitative research** to test the hypotheses we formulated and provide answers for their research questions. Quantitative research is basically **statistical** or numerical research and it works by assigning numbers to the terms we are studying. In our case of the Democratic Peace

Proposition, when the independent variable took the value "democratic", it got a "1" for example, and when it took the value "non-democratic" it got a "2." The researchers also assigned numbers for the values of the dependent variable. When it assumed the value "peaceful crisis behavior" or "settlement" it got a "3" and when it took the value "conflictual crisis behavior" or "war" it got a "4." Afterwards, with the aid of computers, the researchers determined whether the "1s" tended to match up with the "3s."

Please notice that we very much simplified the actual process the quantitative researchers were engaged in. However, in essence, quantitative research is about establishing relationships, patterns, or **correlations** between two or more variables. In our specific case they attempted to determine whether there was a correlation between the type of regime **(IV)** on the one hand and crisis behavior or crisis outcome on the other hand **(DV)**. Another word for correlation is **law** and we distinguish between **absolute law** (sometimes also referred to as deterministic law) and **probabilistic law**. An absolute law states: "Whenever we observe x, we will observe y" or "If x, then always y." A probabilistic law states: "Whenever we observe x, we will observe y, with the probability z." Almost all conclusions in political science research are made with a particular degree of confidence or certainty, reflecting that they are probabilistic assertions not absolute ones.

Finding the Answer

The results of this quantitative research suggested that while democracies did engage in war with non-democracies, they resolved any crises they had with each other peacefully. In other words, the researchers did not find much support for the monadic version of the hypothesis, but they found strong support for the dyadic version of the hypothesis. In fact, the strong support for the dyadic version has been described to approach the ideal of an **absolute law** or a perfect correlation. In other words: Democracies never go to war with each other. This is a strong conclusion and we shall return and discuss it critically. First, however, we need to address a few more things about the research process.

Quantitative researchers addressed the *whether* question, that is *whether* democracies are more peaceful generally or whether they are peaceful more particularly among themselves. What was missing was an answer to the *why* question – the **causal mechanism**. A causal mechanism is not about *that* we observe a certain relationship (this is what laws are about); it is about *why* we observe this relationship, that is, the *reason* behind it. In other words, it is the explanation for *why* something happens. In our case: Why are democracies not fighting each other? Answering this question became the task for **qualitative research**. In simple terms, qualitative research is non-numerical, in-depth research or data collection limited to a few cases. There are many types of qualitative research. Among them are, for example: observation, field research, interviews, document analysis, case studies and others. Some of the most prominent qualitative methods in political science will be discussed in subsequent chapters of this book.

The distinction about *whether*-questions and *why*-questions illustrates nicely that **quantitative** and **qualitative** research methods can complement each other. It may surprise you to hear that political scientists often argue over which method is the better one. **Quantitative** researchers sometimes claim that statistical research is more systematic and objective. **Qualitative** researchers argue that their research is more context-sensitive and empathetic. Indeed, both methods have separate, but complementary advantages. **Quantitative methods** are well suited to test **hypotheses** and deliver correlational conclusions about purported causes and effects. **Qualitative methods**, on the other hand, are well suited to investigate the underlying causal mechanisms of established relationships. They should be cast not as competitors, but as collaborators towards addressing the same research questions.

Let us now return to our discussion of the Democratic Peace Theory and the efforts of qualitative researchers. This group of scholars studied in depth and in very great detail the cases where two democracies were in a crisis with each other, but where they ultimately settled a disagreement peacefully. They found that in these crises there were often inclinations for war, sometimes very strong inclinations. But, fortunately, the situation did not come to it. They identified two causal mechanisms to explain this outcome. One mechanism was labeled the "cultural argument" and the other mechanism was labeled the "structural

argument." Let us look at each causal mechanism in more detail.

According to the "cultural" argument, a democratic political culture is distinct from other types of political cultures such as authoritarianism. It is said that democratic rules and regulations create a culture that socializes the members of the polity into norms of mutual tolerance, collaboration and compromise, and encourages peaceful means of internal conflict resolution. It is also said that this culture comes to apply, not only within a democracy, but also across national boundaries toward other democracies. Here we encounter yet another type of **variable** in addition to the ones we discussed already, namely an **intervening variable**. An intervening variable is an intermediate factor between two other variables (IV and DV) in a causal chain – it is caused by the independent variable and it causes the dependent variable.

It will be evident to you what the **intervening variable** in the present case is. It is the culture of conflict resolution as we just described it. This culture is caused by the democratic governance of the state and it causes democratic states to approach each other in like manner – in an effort to mediate the crisis and any disagreement, rather than to prevail over the other through the use of force.

The "structural" argument is different, but also somewhat related. According to the structural argument it is the internal makeup (the structures) of democracies that are responsible for the observed peace among democracies. The architecture of democracies is fundamentally different from non-democracies. The main characteristic of a democracy is that it is constituted through a system of checks and balances. Who or what are the "checkers" and "balancers"? Some of them are quite obvious. For example, the three independent branches of the United States government, namely the Executive, the Judiciary and the Legislative act as checks and balances on each other. But there are also other checkers and balancers such as the media and the people, for example. In short, in a democracy, no single leader, group or institution can simply declare war. It requires a broad consensus, and this consensus is hard to come by because war is something that is costly in many ways. The checkers and balancers act as brakes on further escalation and allow more time for peaceful conflict resolution. In an authoritarian state, by contrast, if the dictator wants to go to war, there is not too much that can stop him. There are no checkers and balancers that could put on the brakes on an escalation course and avert war.

Questioning the Answer

We have stated that researchers found that democracies indeed do not fight each other, and some went even so far as to claim that this finding is approximating an **absolute law** or a perfect correlation. This is indeed a very strong conclusion. However, just because there *seems* to be good evidence for such a conclusion, does not mean that we should readily accept it. When engaging with theories, propositions, hypotheses, data analysis and conclusions that follow from such, it is very important to remain critical and not to accept them too readily. Let us problematize at least two aspects of the Democratic Peace Theory.

The first aspect we shall problematize is the purported **causality** in the Democratic Peace Theory. In order to establish causality three criteria must be met. First, the assumed independent variable and the dependent variable must correlate. Second, the independent variable must precede the dependent variable in time. These two criteria are generally relatively easy to establish. Also in the present case we can easily see how they are met. The third criterion is that the relationship between the independent variable and the dependent variable must not be **spurious**. A relationship is spurious if the assumed independent variable and the dependent variable are in fact not causally related: if there is, in fact, a third (hidden) **extraneous variable** causing the correlation between them. An **extraneous variable** (sometimes also referred to as a confounding variable) is a factor that could have created an "accidental" or "non-causal" relationship between the independent and dependent variables. The third criterion for causality is generally the most difficult one to establish.

As critical students of politics we should ask: Could the strong relationship between the purported independent variable "regime type" and the dependent variable "crisis behavior/crisis outcome" be **spurious**? In other words, we must ask: Is there a variable that causes both democracy and peaceful crisis behavior or resolution? You will remember that we said that a liberal and open market economy is an antecedent variable to democracy. However, it could also be a confounding variable. How so? The answer is simple: Open market economies often lead to subsequent democratization processes in a country. At the same time, however, it may be well known to you that countries with open market economies tend to engage in commerce and trade with each other and that they,

therefore, have normal (not conflictual) relationships with each other. Thus, in this case, while there might be a relation between democracies and peace, they may not be causally related to each other. Instead, they may both be caused by a third factor – open market economies.

The second aspect we shall problematize is about the operationalization of the Democratic Peace Theory's **concepts**. We already mentioned that the definition of **concepts** such as "democracy" and "war" is very important. In simple terms, what counts as a war? Or in more blunt terms, how many people have to die for a conflict to be counted as a war? If 100 people die, does this qualify as a war? The answer is "no." It will be seen as unfortunate, but it will not be seen as a war. Researchers agreed that a war is a situation in which there are at least 1,000 battle deaths. There is nothing special about this number – it has just become the convention. However, settling on this operationalization allows for the exclusion of some troublesome cases that would cast doubt on the validity of the Democratic Peace Theory. For example, although proponents of the Democratic Peace Theory consider Finland a democracy, the country's alliance with Nazi Germany in the Second World War is dismissed because there were fewer than 1,000 Finnish casualties. Thus, the critics contend that democratic peace theorists tinker with the operationalizations of concepts to dismiss important cases that the theory fails to account for.

Thus, we must ask: Could it be that the conclusion of the Democratic Peace Theory as approximating an absolute law is overstated? Does the theory "work" because it is "set up" to work? We do not want to suggest that the Democratic Peace Theory is wrong. Rather, we want to suggest that we, whether we are students of politics or not, must always remain critical toward theories and any research findings and conclusions they generate. This will allow us to work towards knowledge that we can have more confidence in. This is very important because research should have practical **implications** and **policy relevance**. In other words, the knowledge that we are gaining from our research should benefit us in the real world.

The Implications and Policy Relevance

If it is indeed true that democracies do not fight each other, then the normative implication is to work towards democratization in countries that are lacking this system. Indeed, the Democratic Peace Theory has been applied in the real world. More specifically, it served as a very important guide for the making of foreign policy for various US Presidents. In contemporary times it was used perhaps most explicitly during the administration of President Bill Clinton, but also during the administrations of following Presidents George W. Bush and Barack Obama. You may know that each administration develops at least one National Security Strategy. Usually this is a document of about 30 to 50 pages and it summarizes the main goals of the administration as well as the strategies by which these goals should be attained. The National Security Strategy of the Clinton administration was titled *A National Security Strategy of Engagement and Enlargement.*

You might ask "engagement and enlargement of what?" The answer is engagement of the United States in the world to enlarge the zone of democracies. In simple language, the task for the US was to encourage and support the inception, building and furthering of democracies. This can happen through various means, such as financial and economic aid, trade agreements, support of democratic grassroots movements in the target country, (military) intervention and through many other ways. Whatever means are chosen towards the enlargement of the zone of democracies, the endeavor is often very costly. Yet, the makers of US foreign policy have been willing to take on the cost exactly because they believe that the Democratic Peace Theory is true. In the Preface to the *National Security Strategy of Engagement and Enlargement*, President Bill Clinton wrote, "The community of democratic nations is growing, enhancing the prospects for political stability [and] peaceful conflict resolution." On the second page, the document argues "we know that the larger the pool of democracies, the better off we, and the entire community of nations will be." And on the fifth page it expressed a "firm commitment [of the United States] to expanding the global realm of democracy." In fact, leaders in Washington have taken action. The document points out that the United States

has supported South Africa's ... transformation, provided aid to a new democratic Russia and other new independent states as well as Central and Eastern European

nations, assisted Cambodia, and worked with our Western Hemisphere neighbors restoring the democratically elected government in Haiti and hosting the Summit of the Americas, which reaffirmed and strengthened our mutual commitment to democracy.

These are examples from the 1990s. More recent examples of US engagement toward building or furthering democracies are to be found in Afghanistan and Iraq. It is the explicit goal of the US to further the establishment of democracies and strengthen those that have been established already. Of course, these efforts do not seem to be very successful. Both countries are far from being democracies, and in both countries we see extremism, chaos and continuous conflict. US interventions and engagement have not brought forth democracies. In fact, some scholars argue that the US has not only failed with this ambition in Afghanistan and Iraq, but that it has instead upset much of the region and led to ongoing strife.

Is this assertion indeed true, or, at least somewhat true? If so, where does this lead us? Might it be that the Democratic Peace Theory is valid, but that policy implications derived from it are very difficult, or even dangerous to pursue? Should the US abstain from efforts of nation-building? These are the very important questions that now pose themselves. And this is now where your theoretically informed policy conversation can and perhaps should continue.

Exercises and Discussion Questions:

1. Do you think that the Democratic Peace Theory leads to good foreign policy?
2. Generally speaking, do you believe that politicians can learn from political scientists?
3. Provide examples for absolute laws and probabilistic laws. These do not have to be political examples. Subsequently, formulate the corresponding theories.

4. Do you believe the Democratic Peace Theory is valid/true? In what other ways than the ones discussed above, could the Democratic Peace Theory be criticized?
5. If researchers found that negative campaign ads lead to lower voter turnouts in national elections, what would the implications be?
6. What is the possible ramification if war were operationalized as a situation where there are not 1,000 battle deaths, but 100?

Recommended Resources:

For readers interested in learning more about the Democratic Peace Theory, we recommend the following books:

Elman, Miriam, ed. 1997. *Paths to Peace: Is Democracy the Answer?* Cambridge: MIT Press.

Kant, Immanuel. 1975. *To Perpetual Peace: A Philosophical Sketch.* Translated by Ted Humphrey. Indianapolis: Hackett.

Ray, James Lee. 1998. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition.* Columbia: University of South Carolina Press.

For readers interested in learning more about foreign policy and international politics, we recommend the *Foreign Affairs* magazine (foreignaffairs.com) as well as the website foreignpolicy.com

Notes

1 Walt 2014.

2 Kant 1795: 113.

References

Kant, Immanuel. 1795. *To Perpetual Peace: A Philosophical Sketch.* Translated by
   Ted Humphrey. Indianapolis: Hackett.
Walt, Stephen M. 2014. "Do No (More) Harm." *Foreign Policy.* August 07, 2014.
   http://foreignpolicy.com/2014/08/07/do-no-more-harm/.

The Comparative Case Study Method

"Uncivil Society" in the Arab Uprisings

Zaid Eyadat

CONTENTS

Getting Curious: Why do Previous Theories Fail to Explain the Arab Uprisings?

Political scientists got busy as they witnessed a series of revolts in the Middle East and North Africa (MENA) region in 2010. The uprisings were triggered by a Tunisian street vendor, Mohamed Bouazizi. He lived in Sidi Bouzid, a city approximately 260 kilometers away from Tunis. For months the municipal government had humiliated him and denied his basic rights. On December 17, 2010, in his desperation, he set himself on fire and died on January 4, 2011. His self-immolation and death ignited a series of mass protests. Only ten days after his death, the Tunisian government was overthrown and the wave of change spread to its neighboring countries such as Oman, Yemen, Egypt, Syria, Morocco and Libya. In the months to follow, political scientists embarked upon studying what happened and here, as is the case with studying anything, keen observations mattered. For a political science student, there are mainly two things to observe: 1) the political phenomena themselves, and 2) the academic reactions to the phenomena.

Observing the flood of emerging writings, I decided to write myself on the topic, on specific aspects that I had long carried in my mind. My concerns were twofold: 1) scholarly attention to Arab politics had been excessively focusing on persisting authoritarianism, and the role of political Islam, or more specifically, the role of radical Islam in Arab politics, and 2) consequently, the discourse on the role of political parties and people's (in my study, it is labelled as "uncivil society" in contrast to "civil society") political participation in the Arab world remained largely underdeveloped. Therefore, I believed, the ongoing scholarly attempts to analyze the Arab Uprisings of 2010–2011 were missing some crucial factors.

Prior to the Arab Uprisings, democratic transition was often explained with reference to Central and Eastern European as well as Latin American examples. In explaining democratic transitions in these regions, "political parties" have been regarded as the key "players." However, the upheavals in the Arab world have been marked by the absence of these players. At the same time, the ceaseless uprisings in the Arab world demonstrated that authoritarianism may not be as resilient as previously thought.

My point is this: the Arab world has its own unique sociocultural makeup and political history. An explanation of transitory movements in the Middle East cannot afford to bypass these factors. The task for me was to translate my

reflections and observations into a scientific research paper. Before going further, let me state what I had. First, I had political phenomena (reality), namely the Arab Uprisings of 2010–2011. These are a series of events that took place with the intention of political changes in different countries of the MENA (Middle East and North Africa) region. My specific interests about these phenomena were: "Who did it?" (players), "Why did they do it?" (causes), and "What happened as a result?" (outcomes). Second, I had the "scholarly reactions" to the Arab Uprisings. A review of the literature led me to conclude that existing accounts were not satisfactory. The scholarly treatments of the Arab revolutions are missing crucial factors.

The Research Strategy: What is the Comparative Case Study?

Now I needed to develop a research strategy. In simple words, I needed to plan how I was going to conduct my research with the aid of theoretical and methodological tools. A **theory** is a general statement which can describe the causes and/or effects of phenomena.[1] It allows us not only to explain or describe what we observe, but also to predict future occurrences.[2] However, theories are rarely perfect, and they cannot explain the phenomenon you're interested in in a holistic fashion. Instead, theories are "probabilistic, partial, conditional, and provisional."[3] So exceptions to the rule can be found at any time, in any place and under any circumstances.[4]

Nevertheless, we need to explain political phenomena by using and applying existing theories when conducting research. This is because our research needs to be grounded on a plausible, reasonable and logical plane. Also, as political scientists, we find ourselves in a cumulative scholarly discourse. We do not do our research only for ourselves to read and reflect on alone. Rather, we need to regard our writing as a part of academic dialogue in the field of political science. When you put yourself in an academic setting, you need to work like an academic and that happens with the aid of theories. In sum, theories facilitate a "shared language" of the academic world.

While theories are ways of connecting concepts in a reasoned and plausible way, research methodologies, on the other hand, are ways of examining the validity/strength of these relationships. If theories facilitate a "shared language," research methodology is its "grammar." What is a grammar for? As a shared rule, it enables people to compare different languages and helps one to learn a foreign language. Imagine how you possibly would master a foreign language without having any grammatical knowledge. Probably you would have to memorize every single sentence, one by one. It's the same with research methodology. It offers a fixed structure and frame that has been developed, agreed upon and shared by the scholars in the past to facilitate the communication of its users (both researcher and readers). A research method does not benefit only the researcher by offering a "procedural template" for empirical inquiry, it also helps readers to understand others' research better and more quickly.

Which method to use depends on your research question(s). In other words, it is your research question that influences your choice of research method. Theoretically, the conclusion of your research should be the same regardless of the method you used for your research. However, this is not always the case. There are, indeed, several researchers who demonstrate how different research methods can affect their conclusions substantially.[5] Nevertheless, finding a research method that suits your research question best is one of the most important steps in your research.

The research method I selected for my research was the comparative case study. One way to understand this term is to dissect it: **comparative method** + **case study** = comparative case study. The term "comparative method" indicates the way you conduct your research and "case" refers to the "target" of your comparison. Thus conducting a comparative case study means that you study cases in a systematically comparative manner. According to Yin, there are three conditions that make the case study the best methodological option. These are: 1) when "how" or "why" questions are being asked, 2) when the researcher has little control over phenomenon of interest, and 3) "when the focus is on a contemporary phenomenon within some real-life context."[6] In my study, the countries of Tunisia and Egypt are the cases, and I am going to compare them with the guidance of theories on the role of political parties and uncivil society during the transition to find out why the previous theories fall short in explaining

the transitions in the Arab world.

To understand the comparative case study method better, we may visit its historical development. The comparative method was first used by "the founders of modern sociology" such as Auguste Comte and Herbert Spencer, and it was then systematized by John Stuart Mill.[7] It was later adopted by Max Weber and Emile Durkheim to observe "industrial capitalism, the division of labour, religion, and other social processes."[8] Generally speaking, the foremost goal of the comparative method is to find a "common denominator," a universal element that will contribute to broad explanations of the phenomena we're interested in. In a broad sense, all political science research methods can have a "comparative" label as all research aims at **generalization**.[9]

This leads us into a brief discussion about the important task of case selection. Our goal in selecting cases must be to allow us the drawing of causal inferences. Toward this end, there are two types of case-selection approaches, John Stuart Mill's **method of agreement** and his **method of difference**.[10] Both methods have two requirements. The first requirement for the method of agreement is that there is "agreement" on the value of the dependent variable. In other words, in all cases that are being examined, we must have the same outcome on the right side of the equation. The second requirement pertains to the left side of the equation, the side of possible independent variables. Here the selected cases must be dissimilar, except in one regard. For this reason the method of agreement is also referred to as "the least similar" case comparison.[11] The logic here is straightforward: If on the left side of the equation all variables are dissimilar save one, then we can make a causal inference about this particular variable.

An example will illustrate this for you. Figure 3.1 shows two cases and they have the same outcome, namely Y. Comparing Cases 1 and 2, we see that the only common independent variable they share is variable A. Thus, we can infer that A causes Y.

	Independent Variables	Dependent Variables
Case 1	A B C D E F	Y
Case 2	A G H I J K	Y

Figure 3.1 Mill's Method of Agreement

Now to Mill's **method of difference**. This method also has two requirements. First, the outcome in the examined cases must be different. Second, on the left side of the equation, all possible **independent variables** should be similar, with only one variable constituting an exception. Thus, we can infer that this exceptional variable is the one causing the different outcomes for the two cases. The method of difference is also referred to as the "most similar" case comparison because the selected cases should be as identical as possible in all respects except for the value of only one independent variable.

	Independent Variables	Dependent Variables
Case 1	A B C D E F	Y
Case 2	–A B C D E F	–Y

Figure 3.2 Mill's Method of Difference

For example, Figure 3.2 shows two cases which have the different outcomes, namely Y and –Y. Comparing these two cases, we see that they are identical for all but one variable, which takes the value of A (present) in Case 1 and –A (absent) in case 2. Thus, our inference is that A causes Y while –A causes –Y.

In selecting cases for my research, I applied the method of agreement: both Tunisia and Egypt achieved democratic transitions as a result of uprisings (Ys) and each nation had in common the "active role of uncivil society" (A) while the nations varied on other political and social factors (other independent variables).

Now let's turn from our discussion of the comparative method to a discussion of what a case study is. A 'case' is the unit of study. In the sub-field of comparative politics, a case often refers to a state. However, it also can be other political units such as regions (within states), cities, villages, communities or specific institutions like political parties and interest groups.[12] As you can see from these examples, a case connotes "spatial boundaries" perhaps more than "temporal boundaries."[13] Nevertheless, the latter shall also be borne in mind. In my own research, for example, I do not mean to compare the Tunisian and Egyptian revolts in the nineteenth century. Instead, I research the Arab Uprisings that began in late 2010. Additionally, a case study connotes "intensive study." It implies the systematic in-depth study of one or a few cases, rather than a brief review of multiple cases.[14] Because of the rich insights that one can gain through

case studies, the method has been welcomed in a number of disciplines such as anthropology, archeology, history, education, medicine, psychology, sociology, social work, and of course, political science.[15] The literature suggests five situations in which the case study method is especially appropriate:

1. When the researcher wants to establish a theory or theories;
2. When testing theories that already exist;
3. When identifying a previous condition or conditions that lead or contribute to a phenomenon;
4. When the researcher wants to establish the relative importance of those contributing conditions; and
5. When trying to establish the fundamental importance of the case with regard to other potential examples.[16]

Issues with this Research Strategy

The chief merit of the case study method is the in-depth analysis that it offers. The case studies that humankind owes the most to might well be Charles Darwin's study on evolution after his visits to some select locations on Easter Island, Sigmund Freud's quest on human psychology based on less than a dozen clinical cases, and Jean Piaget's research on human cognitive development from the observation of his own two children.[17]

Yet, the case study method has also received harsh criticisms, mainly for two reasons: 1) it is said to lack generalizability, and 2) it is said to suffer from selection bias. The best example of the first criticism might be an in-depth study of the French Revolution. It can tell us something about revolutions in general, but then again, fails to apply to other revolutions, including the Arab revolutions. The prescription for this problem is to increase the number of cases, which will only make a research study an endless task.[18] If someone criticizes my research with this issue, the question would be: "Can Tunisia and Egypt represent the other 20 countries of the MENA region?" For the same reason, the comparative

method has also been criticized as it does not include enough cases to generate generalizable conclusions.[19]

Meanwhile, the criticism of selection bias means the researcher's alleged lack of objectivity in selecting case(s). The researcher may feel inclined to select those cases for study that are more likely to satisfy his hypothesis. Of course, as researchers we should guard ourselves against such temptations.

However, there have been a number of scholars who have rigorously defended the case study method against such criticisms. To the criticism that a case study is a "mere" case study for the small number of cases and thereby methodologically flawed, the defenders counter that the focus should be put on the intensiveness that a case study offers.[20] For the problem of lacking generalizability or representativeness, Robert Yin recognizes the academic contribution of case studies as they are indeed "generalizable to theoretical propositions" although not to entire populations or universes as the researcher's goal is to expand previous theories or generalize hypotheses.[21]

The Literature Review

After you have a research question and research method in your mind, you need to write a literature review. A literature review is a section where you critically engage with what other scholars have written about your research topic. In this section, what you do is not simply narrate what other scholars or representative scholars have written on the issue. Instead, you need to review how the related discourse has been created and developed, and how it has evolved in academia. For instance, you can organize the literature chronologically, by schools of thought, or by specific themes. I have seen many undergraduate students have some anxiety when it comes to conducting their literature review. If you are one of them, do not panic. Try to think that this is a chance for you to review what others have researched about your topic. You will learn a lot about the topic itself and it will be of benefit to your research and academic growth. Yin advised students to regard a literature review as "a means to an end, and not – as most students first think – an end in itself."[22] He added, "Budding investigators think

that the purpose of a literature review is to determine the *answers* about what is known on a topic; in contrast, experienced investigators review previous research to develop sharper and more insightful *questions* about the topic."[23]

Another question that I often get from students is about the location of the literature review. Many students wonder whether to include the literature review in the introduction or rather to make it a separate section and, if the latter, exactly where. The answer to this question is that there is no fixed rule on the location of literature review. Conventionally, however, it is placed at the earlier stage of your writing, especially before any empirical, theoretical or methodological section. This way your literature review can be a stepping stone for your argument to be elaborated on. Also, doing so will help your readers to catch the academic position of your idea more easily than when they find it at the later stage. In terms of its labeling, and whether to devote a separate section subtitled "Literature Review" or not depends on the type of writing. For instance, if you are writing a master's or a doctoral thesis, it is recommended to have a separate section subtitled "Literature Review." However, for an academic journal, you do not necessarily need to have a section that is subtitled "Literature Review." Rather, you can put a section with a subtitle which can capture the key aspects of your review.

In my case, I wanted to discuss the role of political parties in the Arab world and their role in transitions in Egypt and Tunisia during the Arab revolutions of 2010−2011. The first thing I did was review literature on the definition of political parties, and theories on the role of political parties in democratic transition. "Reviewing definitions" might sound challenging and pointless, as most students tend to believe that what matters most is a "new and critical argument." Yet, we must remember that every argument is composed of concepts. Each of them has its own history of development and their interpretations can differ among scholars. For example, there is no single definition of "democracy." Rather, there are only "general scholarly agreements" on specific conceptual elements that construct the definition of democracy.[24] As such, the first step in developing my arguments was to talk about the definition of political party.

Indeed, scholars have given many definitions of political parties. James Kenneth White points out that a political party denotes "any group, however loosely organized, seeking to elect government officeholders under a given

label."[25] Joseph A. Schlesinger states that "a political party is a group organized to gain control of the government in the name of the group by winning election to public office."[26] John H. Aldrich, on the other hand, asserts that "political parties can be seen as coalitions of elites to capture and use political office. [But] a political party is more than a coalition. A political party is an institutionalized coalition, one that has adopted rules, norms and procedures."[27]

As such, the definitions of political parties differ, with varying emphasis on the elements that construct each definition, such as "public office," "winning election" or "coalitions of elites." However, the review confirms that most scholars are in agreement regarding the inevitability and necessity of political parties "in regime formation processes" although they do not agree on the definition of political parties and how they operate.[28] Thus, political parties have the ability to join and mobilize individuals into large groups that are capable of impacting significant political outcomes. While those outcomes are diverse, political parties so far have been regarded as "effective instruments," hence the actual players, for either the continuation of authoritarian regimes or playing "key roles in moving a polity toward democracy."[29]

Along with the above review on the definition of political parties and their standing in the MENA region, I needed to review further the existing literature on the role of political parties in transition movements. Here I noted that some of the most important sources of democratization theory have dealt with Latin American and Eastern European transitional processes. Democratization occurred in both of these regions during the "third wave" of democracy that began in the early 1970s. This wave, as claimed by Samuel P. Huntington, was triggered by challenges to authoritarian legitimacy, global economic growth, changes of religious institutions, push for human rights and democracy by external actors, and the demonstration effects of democratization in other countries.[30]

Over the years, many scholars have assessed the key factors of transitions in those regions. Their work can be divided into two groups: macro-and micro-level analyses. Macro-level theories predominantly stress socioeconomic conditions and cultural compositions for the political transition process. To be more specific, these theories regard economic development and modernization as a main driver of democratic transition which facilitate the ability of the masses to launch and sustain collective action for common political demands and for pressuring the

government and its authorities effectively.[31] However, these theories are not sufficient in explaining democratic transition processes as there are often real-world cases that prove the significance of a different set of factors. In micro-level model theories, democratic transitions are seen as a series of strategic interactions, often among subgroups of society. Scholars in this group, like Munck, have used the Latin American cases, where experienced democratic transition soon regressed to autocracy, re-democratized, and then struggled for the survival of democracy.[32] In these analyses, more focus was put on actors, their contexts, activities and strategies than on economic development and modernization. Yet, neither the macro-nor the micro-level theories could explain the 2010–2011 Arab revolutions properly. This time in the Arab region, the main actors had no political titles or positions, thereby being at odds with previous studies.

Conducting the Study: Collecting Data and Performing the Analysis

Having reviewed the literature on political parties and transition theories based on the historical contexts of Central and Eastern Europe and Latin America, the next step in my study was to analyze how the theories have proven to have little applicability for the Arab Uprisings. Specifically, it was necessary to analyze the factors that have prevented political parties from having a major impact on the course of the revolutions. For this purpose, I need to draw on what has been said and predicted about Arab politics and compare this to what really happened in Tunisia and Egypt during the Arab revolutions. Here I shall lay out some of the main insights I have gained by looking at the explanations and interpretations which were provided by academia and media following the Arab revolutions.

One of the major differences between the 2010 and 2011 Arab revolutions and previous democratic transitions is that in the Arab revolutions, organized civil society groups, including political parties, played only a limited role. Prior to the Arab revolutions, some Middle East analysts had assumed that civil society organizations might work as a "magic bullet" to reform the autocratic regimes in

the Arab world, and Western aid money had indeed flown to civil society organizations. However, the role of civil society groups, including Islamists, was negligible during the anti-regime demonstrations in Tunisia and Egypt this time. Moreover, these groups were as surprised as the regimes by the anti-regime protests, demonstrating their weak connection to the street.[33]

"No party, no union, no politician" was involved in Tunisia in forming public mass uprisings.[34] Thus, the regime failed in quelling the revolts due to their unexpected nature and uncatchable structure.[35] It was not very different in Egypt. The Muslim Brotherhood was the most influential civil society group, yet, in the opening days it stayed out of the mass protests until its leaders were pressured to participate by its younger members.[36] A major stream of a particular Salafist doctrine holds that pious Muslims should refrain from involvement in politics to avoid excessive engagement with a sinful material world.[37] Some Salafist leaders initially condemned the demonstrations, because traditional Salafist doctrine holds that the populace must obey rulers as long as they are Muslim in order to prevent *fitna* (strife and disaster).[38] Secular opposition parties in both countries also found themselves marginalized in the revolutions, because many Tunisians and Egyptians viewed them as collaborators of the old regimes and they saw the party leaders as a part of the political elite.[39]

Mainstream theories of transition state that civil society groups will tend to play a significant role in democratization, yet in Egypt and Tunisia participation of the institutionalized civil society such as unions and political parties were limited to the middle class. Unlike the pro-democracy movements in Communist Eastern Europe, no centralized, hierarchical leadership led the uprisings in Tunisia and Egypt.[40] No charismatic figures emerged to lead the protests.[41] Whenever participation of some political party and civil society members was seen, it was generally in their personal capacities rather than as a representing member of the group.[42] Individual citizens used the internet and social media to connect with one another, distribute information, and arrange demonstrations without the presence of specific political leaders.[43] Activists worked "outside traditional party affiliations and outside long-established civil society groups, privileging the creation of their own ad hoc committees with variable membership."[44]

The regimes in Tunisia and Egypt could not cope with these tactics because there was no coherent structure for them to attack.[45] As such, most of the work to overthrow the regimes was carried out by the youth underclass taking to the streets. It was not civil society, but an "uncivil society" of people physically occupying public spaces that drove the uprisings. Some of these socioeconomically disenfranchised youngsters had turned to Salafism and other ideologies in opposition to institutionalized politics.[46] These citizens had long been invisible to the states in which they lived and had been ignored by the state institutions. The social contract is a pact between the citizen and the state. For disenfranchised citizens who have been ignored by the state, acting outside of its institutions, rather than engaging with them, is a way they can exercise their citizenship.

The visible activity of political parties appeared only as the old regimes ended. After the overthrow of the Tunisian government under Zine El Abidine Ben Ali on January 14, 2011 and the Egyptian government under Hosni Mubarak on February 11 of the same year, more than 100 political parties emerged in Tunisia and over 60 in Egypt.[47] While this appeared to indicate a boom for the development of political parties, "the vast majority of these new parties failed to do more than register with the government."[48] In Tunisia, pro-secular forces were fragmented into many small parties.[49] The public mistrusted those political parties that had existed before the revolutions for their connection with the regimes.[50] In exchange for working within authoritarian systems, they were allowed token representation in parliament.[51] They, in effect, provided a fig leaf of democratic legitimacy for the Ben Ali and Mubarak governments by allowing the dictators to claim that they had instituted a multiparty system.[52]

In contrast, the Islamist oppositions in Tunisia and Egypt had faced decades of state repression while in opposition, earning them the respect of much of their populations and making them appear "credible parties of government" to a wide section of voters.[53] Because of their charitable networks and the religious education they offered within communities, the Islamists were also the only political force with deep ties to the grassroots.[54] The political playing field was thus essentially cleared for them in the free elections that followed the revolutions.

In Tunisia's 2011 parliamentary elections, the Islamist Ennahda won nearly 40 percent of the vote and received 89 out of the 217 seats in parliament.[55] Ennahda attracted voters from outside the Islamist sector, "because it was untainted by the corruption that has sullied politics for so long."[56] The secular Congress for the Republic (CPR) party finished a distant second, with only 8.71 percent of the vote and 29 parliamentary seats.[57] The Democratic Forum for Labour and Liberties (Ettakatol), a social-democratic party, attracted the third greatest number of votes and won just over 7 percent of the total and 20 parliamentary seats.[58] Ennahda formed a coalition government with these two parties.[59] However, turnout for the election was low, with only 51.7 percent of eligible voters casting their ballots, indicating a lack of public enthusiasm for the electoral choices available to them.[60] An effective secular opposition in Tunisia eventually emerged in the form of the Nidaa Tunes Party, led by the former prime minister and Ben Ali regime member, Beji Caid Essebsi.[61]

The Islamist victory in Egypt's 2011–2012 elections for the People's Assembly was even more comprehensive, with the Muslim Brotherhood-affiliated Freedom and Justice Party (FJP) and the Salafi Al Nour Party winning a combined 70 percent of the seats.[62] The strong performance of the Salafists is a noteworthy example of political party development in the Arab world. In its drive for political power, the Al Nour party has belied its ultraconservative reputation and demonstrated moderation and willingness to compromise. Al Nour endorsed a non-Salafist position for the 2012 presidential election, and once in parliament it reached out to secular parties to try to prevent a Muslim Brotherhood majority.[63] Although Al Nour did not support the June 30–July 3, 2013 protests against the Muslim Brotherhood-led government of Muhammad Morsi that culminated in a military coup d'état, it supported the military's political "roadmap."[64]

The results of the parliamentary elections, as well as the presidential election in Egypt 2011–2012 won by the FJP candidate Muhammad Morsi, demonstrated an initial electoral dominance of the Islamist parties and an overall weakness of political parties in Egypt. The highest-placed secular party in the legislative elections, the New Wafd, received only 9.2 percent of the vote.[65] The main challengers to Morsi in the presidential election ran as independents.[66]

One legacy of the Mubarak era is that many Egyptians do not consider political

parties to be useful vehicles for change. Under Mubarak, there were 24 licensed political parties, yet in the 2005 parliamentary elections they won only nine seats.[67] Opposition figures such as Muhammad El Baradei wanted to appeal to as broad a segment of the Egyptian population as possible and felt that identifying with a party was unnecessarily divisive.[68] One scholar argued that even the successful FJP and Al Nour parties were not political parties in the strictest sense, because they "rel[ied] on influential pre-existing networks which had primarily nothing to do with politics, and even less with institutionalized politics."[69] The primary root of the parties' popularity was not in their political activities, but in the social services they provided. Also, personality mattered far more than party affiliation to the average Egyptian voter, and they often voted for candidates based on familial or tribal connections.[70] An Egyptian political science professor observed that "people don't elect representatives on the basis of issues yet—the question of identity is currently the most important."[71]

Despite clear election victories by Islamist political parties in Tunisia and Egypt, they have been unable to make a lasting impact on the transitions in their countries. Instead, just as in the revolutions themselves, groups outside of institutionalized politics have driven events. Ennahda failed to consolidate its control over the government as "street politics" continued to be the most influential force in Tunisia. Salafist groups have demonstrated against and attacked symbols of secularism in Tunisia and stormed the US embassy in Tunis in reaction to an anti-Islam film.[72] Many of the economically disadvantaged youths who had made up the bulk of the protesters have joined these Salafist groups because they feel that neither the secular nor the Islamist establishment represents them.[73] Pro-Ennahda gangs calling themselves "Leagues for the Protection of the Revolution" have engaged in deadly clashes with opposition party supporters.[74] The tragic climax of this political violence occurred in 2013 when two secular opposition politicians, Chokri Belaid and Muhammad Brahmi, were assassinated by Salafist extremists. These assassinations touched off a renewed political crisis and mass protests in Tunisia that forced Ennahda to hand over power to a technocratic government tasked with administering new elections scheduled for late 2014.[75]

In the Egyptian case, the country's military has been the main power-player in

politics, and the direction of Egypt's transition has depended on whom the military decided to support. The January 25 revolution overthrew President Hosni Mubarak and his government but not the overriding influence of the military. Toward the end of Mubarak's rule, there were tensions in his relationship with the generals. These concerned his plans to have his son Gamal, who was not a military man, succeed him. These also concerned his neoliberal economic policies that tended to threaten the dominant role of the military in the Egyptian economy.[76] The mass demonstrations in Cairo's Tahrir Square (Freedom Square) finally convinced the generals that Mubarak had become a liability that had to be removed. They themselves retained the determination, capability and popular legitimacy to maintain their central role in Egyptian politics. From the outset, the Muslim Brotherhood and President Morsi faced significant institutional hurdles to implementing its agenda.

The FJP government struggled to get the state machine to work, and in June 2012, a Supreme Constitutional Court ruling dissolved the People's Assembly.[77] In response, the Supreme Council of the Armed Forces (SCAF) declared itself the supreme power in Egypt.[78] President Morsi countered this move in August 2012 by purging the top brass of the military and issuing decrees affirming his place as commander-in-chief.[79] The FJP government, along with its Salafist allies, initiated an uneasy partnership with the SCAF in the drafting of a religious-tinged constitution, and succeeded in excluding secular forces from the process.[80]

At this point it appeared that the transition in Egypt was beginning to follow a traditional transitional model with the Muslim Brotherhood playing the role of an opposition party that successfully takes power. However, the results of the referendum on the new constitution in December 2012 revealed that Egyptians were rapidly withdrawing from participation in institutionalized politics. The constitution was approved with 63.8 percent of the vote, but the turnout was only at 32.9 percent.[81] When Morsi announced that new parliamentary elections were to be held under the new constitution, the National Salvation Front (NSF), a coalition of groups opposed to the Muslim Brotherhood-led government, declared that it would boycott them.[82] The structure of the NSF as well as that of the Tamarod movement that came later was nearly identical to the types of decentralized networks and organizations that participated in the January 25 revolution. These movements were united by the desire to once again overthrow

a government that they hated, but they put forward no coherent political platform or policy agenda of their own.[83] The NSF described itself as "a peaceful, comprehensive alternative outside of institutionalized politics," and kept with its members' attitudes that formal politics in Egypt was a rigged game.[84] The Egyptian opposition thus once again turned to street politics to effect change. In the face of new rounds of massive demonstrations that began on June 30, 2013, the military changed sides and deposed Morsi on July 3, 2013. The secular political parties in Egypt relied on the support of the military to achieve their goals rather than on setting up effective political parties and attempting to defeat the Muslim Brotherhood at the ballot box.[85] In the wake of the overthrow and repression of the Muslim Brotherhood, no significant new political parties emerged to take over management of Egypt's transition; instead, the same military elites that had ruled Egypt since the time of Nasser once again resumed control. The overwhelming favorite in the later presidential election, former army chief and leader of the coup d'état, Abdel Fatteh el-Sisi, was unsurprisingly running as an independent.[86]

The inclination of the secular opposition in both Egypt and Tunisia was to not accept the legitimacy of the Islamist governments even though they had been fairly elected. The NSF repeatedly made demands of the FJP government that ignored the solid Islamist victories in every post-revolution election, including that it should cede power to a national unity government.[87] Similarly, in Tunisia, protesters demanded that the Ennahda-led government transfer power to a technocratic government despite the fact that it controlled more than three times as many seats in parliament as its nearest challenger.[88] Scholars of Middle East politics have to some degree overlooked this lack of respect for democratic outcomes by the secular opposition forces in these countries and have instead tended to focus their attention on Islamist parties and whether Islamism is compatible with democratic values.[89] The widespread disdain for political parties among the secular and liberal protesters may have translated into a disdain for multiparty democracy, because, in the given context, it gave a disproportionate voice to groups that they considered threatening to their way of life.

Lessons to Be Learned

There are two things I learned while doing this research. The first is the importance of observing political phenomena. Critical thinking is also important. Yet, observation takes place earlier than anything. Therefore I would like to emphasize that "good observation matters." When you observe well, you can have meaningful critical thinking, which will lead you to a good research question that is crucial for the success of your research. The second lesson is the importance of theoretical preparedness. Observation without theoretical guidance is likely to end up in one's emotional reaction to the issue at hand. The more theoretically prepared, the sharper you can observe. When you observe better, your scholarly reaction to the political phenomena will be quicker and more precise as you will be able to interpret such phenomena better. This might sound difficult to you. However, the initial difficulty of studying theories will soon fade as you get to read and know more about theories. And, gradually, you will find yourself enjoying learning and reading about theories.

When I was studying Arab politics, I had an idea that theories grounded on the Western model of democratization cannot adequately capture Arab politics and its power dynamics. As such, the Arab revolutions provided me with perfect cases to confirm what I had thought and therefore could now scientifically prove. However, it is important not to misunderstand that new political events require a new thinking all the way down. Instead, regard a political event as a chance to revisit what has been studied by other scholars earlier, and reassess their theories by applying them to the current event. In doing so, maximize the use of academic elements that are sharp, and use critical observation, the scientific method and theoretically guided thinking. When these are well trained, transforming your observation into the academic agenda will be easier for you, simultaneously tightening your academic engagement.

Interested to Know More about the Study Discussed in this Chapter?

Eyadat, Zaid. 2015. "A Transition without Players: The Role of Political Parties in the Arab Revolutions." *Democracy and Security* 11(2): 160–175.

Exercises and Discussion Questions:

1. What are the merits of a comparative case study? What are its shortcomings?
2. When can a case study be the best option for your inquiry?
3. How can a case study researcher collect data or materials for his or her research?
4. What are the two case selection methods? Think of a political example for each and draw an illustrative figure for each.

Recommended Resources:

American Political Science Association's (APSA) Qualitative and Multi-method Research Section/The Consortium on Qualitative Research Methods (CQRM) (www.maxwell.syr.edu/moynihan_cqrm.aspx)

Coppedge, Michael. 2012. *Democratization and Research Methods.* New York: Cambridge University Press.

George, Alexander, and Andrew Bennett. 2005. *Case Studies and Theory Development in Social Sciences.* Cambridge: MIT Press.

Gerring, John. 2007. "The Case Study: What It Is and What It Does." In *The Oxford Handbook of Comparative Politics,* ed. Carles Boix and Susan C. Stokes. New York: Oxford University Press, 90–122.

Notes

1 Van Evera 1997: 7–8.

2 Coppedge 2012: 49.

3 Coppedge 2012: 5.

4 Coppedge 2012: 5.

5 Coppedge 2012: 2.

6 Yin 1984: 1.

7 Hantrais 2007: 4.

8 Hantrais 2007: 4.

9 Burnham *et al.* 2008: 69.

10 Mill 1862.

11 George and Bennett 2005.

12 Gerring 2007: 94–95.

13 Gerring 2007: 95.

14 Gerring 2007: 95.

15 Gerring 2007: 90.

16 McNabb 2010: 238.

17 Gerring 2007: 99.

18 Lijphart 1971: 686.

19 Burnham *et al.* 2008: 88.

20 Gerring 2004: 341.

21 Yin 1984: 10.

22 Yin 1984: 9.

23 Yin 1984: 9.

24 Coppedge 2012: 11.

25 White 2006: 5–15.

26 Schlesinger 1991.

27 Aldrich 1995: 19.

28 Angrist 2007.

29 Angrist 2007: 12.

30 Huntington 1991.

31 Lipset 1959; Huntington 1991.

32 Munck 2001.

33 Cavatorta 2012: 77.

34 Cavatorta 2012: 77.

35 Storm 2014: 2.

36 Cavatorta 2012: 77.

37 Wiktorowicz 2006.

38 Hoigilt and Nome 2014: 39, 49.

39 Collombier 2013; Kusche 2013.

40 Cavatorta 2012: 79.

41 Roy 2012.

42 Cavatorta 2012: 77.

43 Cavatorta 2012: 79; Collombier 2013: 2.

44 Cavatorta 2012: 78.

45 Cavatorta 2012: 79.

46 Merone and Cavatorta 2012.

47 Muasher 2013

48 Muasher 2013.

49 Kusche 2013: 114.

50 Kusche 2013: 118.

51 Collombier 2013: 3.

52 Brownlee 2007.

53 Roy 2012: 5–6.

54 Collombier 2013: 1; Kusche 2013: 117.

55 Tunisia-Live 2011.

56 Kusche 2013: 116.

57 Tunisia-Live 2011.

58 Tunisia-Live 2011.

59 Dworkin 2013.

60 Tunisia-Live 2011.

61 Dworkin 2013: 28.

62 Collombier 2013: 1.

63 Hoigilt and Nome 2014: 48.

64 Saad and El Fegiery 2014.

65 El-Din 2012.

66 Collombier 2013: 1.

67 Elshobaki 2010.

68 Collombier 2013: 5, 14.

69 Collombier 2013: 1.

70 Collombier 2013: 12; Hoigilt and Nome 2014: 42.

71 Hoigilt and Nome 2014: 42.

72 Dworkin 2013: 25.

73 Merone and Cavatorta 2012: 6, 14.

74 Merone and Cavatorta 2012: 29.

75 *Daily News Egypt* 2014.

76 Hasanen and Nuruzzaman 2013.

77 Dworkin 2013: 17.

78 Saad and El Fegiery 2014: 4.

79 Saad and El Fegiery 2014: 4.

80 Saad and El Fegiery 2014: 2, 5.

81 Saad and El Fegiery 2014: 5.

82 Dworkin 2013: 20.

83 Dworkin 2013: 20.

84 Hamzway 2013.

85 Saad and El Fegiery 2014: 6.

86 Al Arabiya News 2014.

87 Dworkin 2013: 21.

88 Tunisia-Live 2011.

89 Storm 2014: 2.

References

Al Arabiya News. 2014. "In Egypt's Political Vacuum, el-Sisi Looms Large."
Associated Press. March 28. www.timesofisrael.com/in-egypts-political-
vacuum-el-sissi-looms-large/.

Aldrich, John H. 1995. *Why Parties? The Origin and Transformation of Political
Parties in America.* Chicago: University of Chicago Press.

Angrist, Michele Penner. 2007. *Party Building in the Modern Middle East.* Seattle:
University of Washington Press.

Brownlee, James. 2007. *Authoritarianism in an Age of Democratization.*
Cambridge: Cambridge University Press.

Burnham, Peter, Karin G. Lutz, Wyn Grant, and Zig Layton-Henry. 2008.
Research Methods in Politics. 2nd edn. New York: Palgrave Macmillan.

Cavatorta, Francesco. 2012. "Arab Spring: The Awakening of Civil Society. A
General Overview." *European Institute of the Mediterranean Yearbook*, 75–81.

Collombier, Virginie. 2013. "Politics without Parties. Political Change and
Democracy: Building in Egypt Before and After the Revolution." European
University Institute (EUI) Working Papers: Max Weber Programme for
Doctoral Studies (MWP).

Coppedge, Michael. 2012. *Democratization and Research Methods.* New York:
Cambridge University Press.

Daily News Egypt. "Tunisia's New Government of Independents Sworn In."
January 29, 2014. https://dailynewsegypt.com/2014/01/29/tunisias-new-

government-independents-sworn/.

Dworkin, Anthony. 2013. *The Struggle for Pluralism after the North African Revolutions.* London: European Council on Foreign Relations.

El-Din, Gamal Essem. 2012. "Egypt's Post-Mubarak Legislative Life Begins amid Tension and Divisions." *Ahram Online,* January 23. http://english.ahram.org.eg/NewsContent/33/100/32384/Elections-/News/Egypts-postMubarak-legislative-lifebegins-amid-te.aspx.

Elshobaki, Amr. 2010. "Parties, Movements, and Prospects for Change in Egypt." *Carnegie Endowment for International Peace,* May 20. http://carnegieendowment.org/sada/?fa=40825.

George, Alexander, and Andrew Bennett. 2005. *Case Studies and Theory Development in Social Sciences.* Cambridge: MIT Press.

Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American Political Science Review* 98(2): 341–354.

Gerring, John. 2007. "The Case Study: What It Is and What It Does." In *The Oxford Handbook of Comparative Politics,* ed. Carles Boix and Susan C. Stokes. New York: Oxford University Press, 90–122.

Hamzawy, Amr. 2013. "A Starting Point: Boycotting as a Campaign and Comprehensive Alternative." The Atlantic Council (March 5). Available at www.atlanticcouncil.org/blogs/menasource/a-starting-point-boycotting-as-a-campaign-and-comprehensive-alternative/.

Hantrains, Linda. 2007. "Contextualization in Cross-National Comparative Research." In *Cross-National Research Methodology & Practice*, ed. Linda Hantrais and Steen Mangen. Cornwall: Routledge, 3–18.

Hasanen, Mohammed M., and Mohammad Nuruzzaman. 2013. "The New Egypt: Sociopolitical Dynamics and the Prospects of the Transition to Democracy." *Mediterranean Journal of Social Sciences* 4(4): 137–145.

Hoigilt, Jacob, and Frida Nome. 2014. "Egyptian Salafism in Revolution." *Journal of Islamic Studies* 25(1): 33–54.

Huntington, Samuel P. 1991. "Democracy's Third Wave." *Journal of Democracy* 2(2): 12–34.

Kusche, Hanna. 2013. "Tunisian Elections after the Arab Spring: A Victory for Democracy or Islamism?" In *The Political Geography of Religious Radicalism: A Compendium of Selected Case Studies from Around the Globe,* ed. Rainer

Rothfuss and Joseph Yakubu. Tübingen: University of Tübingen.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65(3): 682–693.

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1): 69–105.

McNabb, David E. 2010. *Research Methods for Political Science.* 2nd edn. New York: Routledge.

Merone, Fabio, and Francesco Cavatorta. 2012. "Salafist Mouvance and Sheikhism in the Tunisian Democratic Transition." *Working Papers in International Studies*, Center for International Studies, Dublin City University 7: 1–17.

Mill, John Stuart. 1862. *A System of Logic.* London: Parker.

Muasher, Marwan. 2013. "The Path to Sustainable Political Parties in the Arab World." *Carnegie Endowment for International Peace*, November 13, http://carnegieendowment.org/2013/11/13/path-to-sustainable-political-parties-in-arab-world.

Munck, Gerardo L. 2001. "The Regime Question: Theory Building in Democracy Studies." *World Politics* 54(1): 119–144.

Roy, Olivier. 2012. "The Transformation of the Arab World." *Journal of Democracy* 23(3): 5–18.

Saad, Ragab, and Moataz El Fegiery. 2014. "Citizenship in Post-Awakening Egypt: Power Shifts and Conflicting Perceptions." *EUSpring,* January: 1–15.

Schlesinger, Joseph A. 1991. *Political Parties and the Winning of Office.* Ann Arbor: University of Michigan Press.

Storm, Lise. 2014. *Party Politics and the Prospects for Democracy in North Africa.* Boulder, CO: Lynne Rienner.

Tunisia-Live. 2011. "Final Results of Tunisia Election Announced." November 14, www.tunisia-live.net/2011/11/14/tunisian-election-final-results-tables.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science.* Ithaca, NY: Cornell University Press.

White, James Kenneth. 2006. "What Is a Political Party?" In *Handbook of Party Politics,* ed. Richard S. Katz and William J. Crotty. Thousand Oaks, CA: Sage.

Wiktorowicz, Quintan. 2006. "Anatomy of the Salafi Movement." *Studies in Conflict and Terrorism* 29(3): 207–239.

Yin, Robert K. 1984. *Case Study Research Design and Methods.* 2nd edn. Thousand Oaks: Sage.

Field Research

Navigating Politically Sensitive Research in China

Katherine Palmer Kaup

Getting Curious

That doesn't seem fair! Should ethnic minorities in China be held accountable to

state laws if they don't know what those laws are or if they are diligently abiding by a different, but conflicting, set of ethnic laws? What exactly happens when ethnic-minority customary laws conflict with state laws in China, and who gets to decide which laws should be upheld?

I've always been interested in minority rights. It would be hard not to given when I first started studying politics. The year I started doing fieldwork in China, well over a million people died in bloody ethnic violence across the globe. Ethnic conflict is rampant across Asia, with over 800,000 Muslim Rohingyas fleeing from murderous ethnic persecution in Myanmar alone as this book goes to press.

Ethnic tensions in China have erupted throughout the Chinese Communist Party's rule (and long before it, too!). My first book (and the focus of my chapter in this volume's first edition) explored how the Chinese government sought to integrate and manage the Zhuang nationality, China's largest minority group. Through my fieldwork I revealed a story of self-discovery and political mobilization among a diverse collection of ethnic peoples in southwest China, who in response to the Chinese Communist Party's strategic recreation of history came to feel pride in and loyalty to the greater Zhuang ethnic group.

But the Chinese government's interactions with ethnic minorities has been complicated and at times frightening. While the state may have been relatively successful in integrating the Zhuang politically and economically into the People's Republic of China, it has ruthlessly suppressed expressions of ethnic and religious identity among those ethnic groups it fears may want to withdraw from state control. This has been particularly true in CCP policy toward the Uighurs, a Muslim Turkic group in northwest China and toward the Tibetans, a Buddhist minority in western China. In Xinjiang and Tibet, where these groups live respectively, the Party regularly equates expressions of ethnic identity with "The Three Evils" of "ethnic splittism, religious extremism, and international terrorism."[1] The state views Uighur and Tibetan demands for greater religious freedom as a direct threat to its control and as calls to separatism and mobilization against the state. When economics professor Ilham Tohti from Beijing's Central Nationalities University had the audacity to question the wisdom of state policies in Xinjiang, he was sentenced in September 2014 to life in prison for "separatism."[2]

As Senior Advisor on Minority Nationalities Affairs at the Congressional-

Executive Commission (CECC) during my first sabbatical in 2005, I monitored human rights and rule of law developments in minority regions across China. We kept a political prisoner database with detailed information on each prisoner the Chinese government sentenced to jail simply for demanding greater minority rights. I had carefully followed how Beijing strictly enforces its laws and rule in Xinjiang and Tibet, so was curious when I heard of several cases in southwest China in which the government had allowed minorities to use their own ethnic laws, even when they directly contradicted state laws. These were in non-Muslim, non-Tibetan areas, but the contradictions between ethnic and state law were real and apparent. I read about cases in which village leaders were allowed to impose penalties on adulterers, for example, though the national laws explicitly bar non-state actors from imposing penalties for crimes. Why did the government sometimes allow minorities to preserve their customary laws in southwest China? Were there limits to what type of laws the government would tolerate; for example, would the government enforce state laws against sexually harassing women even if minorities did not view those same actions as inappropriate or illegal? What happened when ethnic minority laws and state laws conflicted and who got to decide which laws should rule?

These questions were important, not only to the tens of millions of Chinese citizens who are subject to multiple conflicting laws, but also for what they said about prospects for true rule-of-law development in China. Though a lot of Chinese scholars were writing about these conflicting laws, none thoroughly addressed the question of how these conflicts were resolved. I could find no written guidelines for handling these issues. I had to go to the field to find out for myself.

The Research Strategy: What is Fieldwork Exactly?

So what exactly is **fieldwork**, and why use it? What makes fieldwork unique is that it involves on-site and in-depth gathering of information from human subjects in their everyday settings, where their place in the surroundings informs the study just as much as the fieldworker's questions do.[3] Field researchers

recognize that context matters, often in unexpected and unanticipated ways. By conducting research in a particular location, often for extended periods of time, the field researcher may discover all sorts of variables and factors that impact the research question that she might not have thought to include in a set of survey questions or in an experimental design.

Even when conducting extended interviews in the field, the field researcher may observe that the interviewee's responses are influenced and shaped by factors of which even the interviewee is not aware. I was regularly told, for example, that "minority customary law doesn't actually play much of a role here," only to observe with my own eyes how it shaped land disputes, child custody cases, environmental protection practices, and other issues. Even when asked directly about customary law, the villagers themselves might not have understood its impact: but the context spoke for itself. Villagers' actions were often bound by rules that could be enforced by village religious leaders or elders – these were minority customary laws, even though villagers didn't recognize them as such.

Fieldwork is required to gain an understanding of the complex interplay of social interactions, geography, resource management and other factors that may influence how the people being studied are impacted, and in turn impact, these diverse factors. Fieldwork can produce nuanced findings on sensitive topics generally not available through quantitative methods. In other words, the researcher needs to use fieldwork when written archival materials, surveys, phone interviews or interviews outside of the area alone won't provide the answers needed. Properly conducted fieldwork avoids the hazard of forcing preconceived notions on subjects through survey questionnaires. What is observed in the field is often quite different from what respondents self-report in surveys or interviews.

It is particularly important to conduct fieldwork in environments like China, where information may be strictly controlled or only state-approved versions of reality allowed to be published. Many written sources can only be obtained through personal contacts on site. In China, for example, government offices and research centers gather materials and surveys that are not publicly available. Some of these materials are classified as "*neibu*" or for "internal circulation only," and are legally not allowed to be seen by those without proper clearance.

Scholars who spend enough time in the field to establish contacts who trust them, however, will often be given access to these materials.

"Field research" can incorporate an array of methods, including surveys, interviews, **participant observation**, and collection of written materials for content, textual, or discourse analysis. Surveys ask respondents questions that can be quantified and tabulated. While some surveys might be administered through mailings or over the phone (how many times have you answered the phone, inevitably right in the middle of dinner, to hear a surveyor on the other end?), in developing countries with poor infrastructure, the surveys often have to be administered in person, in the field. Written material can be analyzed in a variety of ways, including looking for how often and in what ways particular topics are covered in the media, government documents, or other outlets: a process known as **content analysis**. Close analysis of text (textual analysis) and the discourse used in speeches or published materials (discourse analysis) can reveal changes in attitudes, or may provide clues as to why a particular policy evolved at a particular time or why a particular event unfolded as it did. Many written materials can only be found in the field, and after reading through them carefully, the scholar may decide to tweak the original research design or pursue new questions that arise from findings revealed by written sources.

Participant observation is a technique in which the researcher observes subjects for extended (though unspecified) periods of time through participating in some role in the community. A scholar studying non-profit organizations' leadership might, for example, contact a battered women's shelter, explain the project to the director, and volunteer to work at the shelter while she observes the organizers' work first hand. Through prolonged observations, the scholar may discover issues, interpersonal dynamics, and challenges that those involved may not even recognize. It is certainly not acceptable to pose as something one is not while clandestinely gathering information on unsuspecting or unwilling subjects, though! When I conducted my fieldwork in Honghe Prefecture, I told people exactly what I was studying. But in addition to my formal interviews, I gathered quite a bit of information through participant observation as I joined a Chinese anthropologist for numerous meals with her family and friends and as I joined her for her fieldwork on a rare Hani minority religious ritual. The latter enabled me to observe the role of the village religious leader in setting, interpreting and

enforcing village laws. I was also able to join in the village religious feast and learn much about village interactions, customs and power hierarchies.

Fieldwork often falls into two broad categories: studies that seek to generate new theories from empirical findings and those seeking to test existing theories and hypothesis through **verification** in the field. Mine fell somewhere in between. Many China scholars are worried that the rule of law is being undermined in China currently as the state seems increasingly willing to apply laws unevenly in the interest of "promoting social harmony" and strengthening Party control. I actually share this concern, though in testing this theory in an ethnic minority field site ultimately concluded that, while greater use of mediation and flexible applications of ethnic and state law may not be an ideal path to rule-of-law development, for the near future it may be a necessary one in minority regions.

There are a number of key issues that any field researcher faces as he or she develops their research strategy. Two of the most important decisions are site selection and deciding on a specific strategy for gathering information. Depending on the research question and the variables being explored, the researcher may need to consider demographics, socioeconomic makeup in the region, geography, transportation access or the political history of the locality. The scholar may be influenced to select a site to further explore a particular conflict, controversy or issue that occurred there, or because the leadership there reflects a certain profile that the scholar wants to understand better. In my case, for this project, I needed to find a site where these conflicts of law were actually occurring. The most likely site, I determined, would be one where there was a high concentration of ethnic minorities, where limited infrastructure would make it less likely that they had much interaction or knowledge of the outside, and where education and income levels were low.

I also needed an "in": some way to get people to talk to me. China is a huge country – I felt a little bit like I was searching for a needle in a haystack. How was I supposed to figure out who had ever experienced or even heard of cases where customary law conflicts with state law? Most cases are not published in China. How could I even figure out what customary law actually entailed for the people bound by it – what were these laws exactly? And then how was I supposed to find judges or police or village heads who were involved in resolving

these conflicts of law, much less find any who would be willing to answer such politically sensitive topics? "Excuse me, judge, but how do you decide when *not* to apply the laws you are legally bound to uphold?" Not a great conversation starter.

I did what any China scholar would do – I looked to my own connections, or "*guanxi*", and expanded out. After receiving official authorization to conduct the study (an arduous process in China, deserving of its own chapter!), I travelled with a well-connected Chinese friend to several potential sites. I spent an entire month exploring which site would be most promising, before even beginning the formal field investigation. I ultimately chose a site that had the features listed above, but was also home to nine different minority groups. That gave me the chance to compare and contrast different customary laws in a single geographic location. I was careful not to choose Muslim or Tibetan minorities, as the Chinese government remains very concerned about foreigners researching these groups, and my access would have been extremely limited. I also chose the home county of a Chinese scholar who agreed to travel with me to introduce me to people and translate several of the minority languages into Mandarin (she was amazing!). Finally, I selected Jinping County in Honghe Prefecture because a major, detailed study on conflicts of minority customary law and state law had been carried out there a few years before I arrived. I therefore knew that these conflicts of law occurred at the site. If I couldn't find further examples, it might be possible that I would end up explaining why my findings were so different from the earlier study. I felt I couldn't lose.

Careful site selection is extremely important and, if not properly done, can alter findings dramatically. To examine the impact of village elections on democratization in China, for example, a scholar obviously needs to go to the countryside. But should the site be in a prosperous region that has had conflict-free elections for decades, or in a remote village in a poverty-stricken county that has just introduced elections within the past few years? Do existing studies suggest that elections are most influenced by levels of education and wealth in a specific area? If so, the scholar might choose a single site with a prosperous, well-educated population to test whether the prevailing theory holds true in China. Or perhaps the scholar will opt to conduct interviews in two different sites, one quite prosperous and one very poor.

Researchers need to think carefully about what other factors might influence or alter findings – does the distance of the site, rich or poor, from the nation's capital matter, for example? How about if one of the sites had a negative experience with elections when they were first introduced, and the others' experiences have consistently been positive? There's a lot to think about! The scholar needs to be as well read and informed about the region as possible to make a good selection. That said, some of the most exciting findings come from completely unexpected places during the field research itself.

In addition to selecting the site/s well, scholars also need to enter the field with a strategy for gathering their information. Thorough preparation before entering the field is crucial: scholars read as widely as possible both theoretical studies that might relate to their research question and empirical studies of the locality where they will be based. I had to learn what I could about each of the nine main minorities in the county I selected for my field site. I delved into anthropological literature on each of the minorities I intended to study. This was an enormously challenging task (and you'll hear more about it again in the "lessons learned" section!). I speak fluent Chinese, but sometimes I wonder if understanding anthropology-ese is even more challenging! Anthropology has its own professional specialized vocabulary, and the studies on China's minorities are rich in anthropological theory, historical references, religious vocabulary, and more. I was trying to extrapolate from broad studies on various minorities' customary practices which could be considered legally binding, trying to do this all in Chinese in very specialized terms, and trying to do it times nine.

Though scholars come to the field with some ideas on how to get started and some working theories on what they may find, a good researcher remains open to new discoveries and unanticipated findings throughout the fieldwork experience. Findings can be distorted and important information overlooked if the research design is overly rigid or built around a set of highly specific and inflexible questions. Researchers will likely use a full toolkit of strategies, including interviews, observation and gathering of written documents.

Those using interviews have to think carefully about how best to design the interviews. There are a number of different types of interviews that can be used. **Structured** and **semi-structured interviews** both raise a series of specific interview questions, usually in a relatively quiet place that has been reserved to

conduct the interviews. In a structured interview, the interviewer raises a specific set of pre-formulated questions, often in a specific order that are usually **closed-ended questions**, or questions that can be answered briefly with a specific answer. For example, "Did your mother ever spank you when you were growing up?" A semi-structured interview allows the interviewer to raise more **open-ended questions** that allow the respondents more room to respond as they would like, and allow the interviewer to ask follow-up questions that arise from the respondent's comments. "What kind of relationship did you have with your parents growing up?" would be an open-ended question.

While the interviewer usually schedules structured or semi-structured interviews, **unstructured interviews** occur more surreptitiously. The researcher may literally bump into someone on the street and strike up a casual conversation. The researcher can then try to gear the conversation towards questions the interviewee finds interesting. Some of my best findings came from unstructured interviews. In the course of interviews and observations, new research questions may emerge or theories that previously seemed irrelevant may prove crucial in explaining the question. The first few days or weeks in the field may be highly exploratory, and depend on open-ended questioning and observation, as well as assessing what access the scholar can garner to explore the question further. Constantly refining both the research question itself and the research design often leads to more exciting and significant conclusions than could be anticipated at the beginning of a project. This is true for all field researchers, but particularly so for those working in developing countries or on politically sensitive issues where access can be difficult and can change quickly.

Issues with this Research Strategy

Fieldwork is often not only beneficial, but absolutely *essential,* when the issue being studied requires subtleties in interpretation or when dealing with a politically sensitive topic that leads to biased or censored reporting and analysis. Surveys often miss nuances needed to understand complex issues such as identity and motivations for collective mobilization. Moreover, though surveys are

increasingly used in China, they cannot be conducted from the States. Survey techniques can be quite different in China, and rely on experienced surveyors conducting the surveys in person in the field.[4] For example, many rural respondents in China have never taken a survey, are distrustful of them and unwilling to put their answers in writing. Many also have trouble interpreting the surveys due to low levels of education. There are a host of other challenges, including insuring that the person or household given the survey actually takes the survey rather than a proxy. Many of these issues can, however, be addressed.

Fieldwork is often the only method of studying the complexity of relationships within a particular site, but it can be difficult to determine how well findings can be generalized. I may have found the flexible application of state and minority customary law, but was that true just in Yunnan? I knew it wasn't true in Xinjiang and Tibet, but what about in the neighboring Guangxi Zhuang Autonomous Region? If the State is willing to allow minorities who have not been fully exposed to state law to utilize their existing customary laws, what is the "tipping" point in terms of how much exposure minorities must have had to state law before the state expects them to adopt state law and cast aside conflicting minority legal practices? By comparing more and more sites, it may be possible to become more precise in identifying key variables that influence patterns of behavior, but it is impossible to study an infinite number of sites. There is an ongoing debate among scholars over whether it is best to use a large number of cases for greater breadth or a smaller number of cases to achieve more depth.

Some field studies focus on a single site, explaining patterns of behavior by providing rich details about the numerous factors that contribute to that behavior. Many researchers opt instead for a comparative study looking at several sites to find differences among them by either selecting the most similar sites or most different sites and controlling for variables. In the most similar site comparative method, scholars identify a variable (level of development, for example) that they think may influence behavior. They then choose two sites that are as closely similar as possible in all ways *except* for level of development (see **method of difference** in [Chapter 3](#)). They may choose two villages that share the same geographical features, have common ethnic demographics and are the same size, for example. This then allows them to argue that level of development,

rather than the other variables, is the cause for any differences they discover in behavior. In Table 4.1, the researcher would choose Villages C and D for the most similar case method field sites. In the most different sites approach, in contrast, the scholar looks for sites that are extremely different *except* for their levels of development (Villages A and B in the chart).

Table 4.1 Comparison of Villages

Village	Development Level (Similar?)	Geographic Traits (Similar?)	Ethnic Demographics (Similar?)	Size of Village (Similar?)
Village A	Yes	No	No	No
Village B	Yes	No	No	No
Village C	No	Yes	Yes	Yes
Village D	No	Yes	Yes	Yes

These comparative approaches can be useful for teasing out the relative importance of a particular variable in influencing behavior, but can be limiting in focusing in on only a few variables and therefore missing nuances that could emerge through a more open-ended design. Selecting multiple sites does not necessarily require comparing across them, but could help explain a single case or question, in what one scholar calls a one-case multi-field-site approach.[5] Selecting more than one field site, though time consuming and sometimes expensive, can help minimize the risk that interview respondents at one site might not prove helpful, or that what one hopes to observe at a particular site in the end is not present.

On a very practical level, one of the primary costs of opting to conduct field research is ... costs. The financial costs of conducting fieldwork can be particularly high if you need to travel long distances to your site, or if you need to be in the field for extended periods of time. Fieldwork in China obviously requires an international flight, hotels and local transportation. In addition, visas

are expensive, and there are a host of required, but often unanticipated costs, including gifts for hosts, meals for interviewees, translation services, *etc.*

Though fieldwork can provide rich and nuanced findings, the method itself can sometimes distort findings if the researcher is not alert to several potential problems. Field researchers must be on the constant lookout for how their own presence and role in the research process might impact their findings. I found this less problematic in this most recent study than in my first project, since the current project was more tailored toward finding specific cases and specific examples of how they were handled rather than seeking to understand how people felt about them. In the first project, I was trying to determine whether ethnicity played an important role in how people labeled "Zhuang" understood themselves, their place in society and their political decision-making. Wouldn't the very presence of a foreign scholar studying the Zhuang lead people to discuss their ethnicity and reflect on it in ways they might not ordinarily? Findings can also be distorted by overreliance on particularly helpful sources, whether those be written sources or a few key contacts who lead the researcher to explore questions in a particular direction.

Researchers also need to be sure that the research methods they are employing don't distort their findings. They must be careful, for example, not to impose theoretical tools and categories inappropriate for the unique context of the field. Again, this issue was more pronounced in my first study when I asked several interviewees if Zhuang had any particular political interests different from Han Chinese. Very few understood the concept of "interests" much less how these interests might be uniquely Zhuang. If I had employed a rigid set of closed-ended questions in structured interviews, I would have erroneously concluded that Zhuang "interests" are not the least important. Instead, I had to use a number of different questions to get at the concept, and avoided using **structured interview** questions that would give respondents only limited options for responding. I also had to play around with vocabulary a bit in this most current project since locals didn't generally understand the term "minority customary law." There are no hard and fast rules or easy fixes for insuring the researcher's role in the gathering of material does not influence or distort findings, but being constantly aware of this possibility is important as you shape your research design and carry out the project.

The Literature Review

I started this research project simply because I wanted answers to a few concrete questions: What happens when minority customary law conflicts with state law? Does the state just impose its own laws, even if the minorities' legal consciousness is different and they define crimes differently? I discovered that in cases of conflicting laws, sometimes state law is applied and sometimes it is set aside in favor of minority customary law. So who gets to decide when and how that's done? I knew the answer to these questions would be important to the millions living under multiple competing legal systems, but I still needed that "so what" answer on how this fit into an even bigger picture and bigger discussion. I needed to engage with what other scholars were saying about these and related issues.

So … first off, other Western scholars generally *weren't* talking about this at all. An excellent recent study by Matthew Erie on Hui customary law was published right after I returned from the field, but very few others were talking about minority laws in China at all. There was quite a bit of literature on how the Regional Ethnic Autonomy Law was being implemented, particularly in Xinjiang and Tibet, but very few studies focused on local utilization of that law or on rule-of-law promotion in minority areas. There is a robust and growing body of English-language literature on rule-of-law promotion in China, but most seem to ignore completely how rule of law might be developing differently in minority regions. That was it! I could fit myself into the broader rule-of-law promotion literature in China, and by focusing on unique challenges in minority regions, I would fill a gap in the English literature.

I couldn't quite stop there, however. China wasn't the only country with ethnic minorities and multiple legal codes. How have other countries handled conflicts between state law and minority customary law? Could China learn from any of them? Since I could find so little on China's minority customary law situation in English, I also had to rely heavily on the Chinese literature. There are literally hundreds of articles and books on minority customary law in Chinese. As I tried to impose some order on the reports and determine what the main research questions and approaches have been (which is essentially the very definition of a literature review!), I realized that providing that story to a non-

Chinese speaking readership was also a contribution to the literature.

Conducting the Study: Collecting the Data and doing the Analysis

Completing the literature review was just one of many steps needed to get to the field. I also had to obtain Chinese government permits for conducting research in China. Depending on their research topic, some China scholars opt not to apply for official research approval. Though some of my shorter trips to China to visit established contacts have proven quite informative for various research projects, I would not venture a full-scale research project requiring extensive interviewing, time in the field or interactions with strangers without official approval. Official approval and the required hosting organization provide the scholar and those whom he or she contacts a degree of protection against government sanctions.

I may spoil the punchline, but let me start first with sharing my conclusions, and then walk through a few of the steps that led me there. The Chinese government is eager to promote the rule of law across the country (with the essential, not to be forgotten, caveat that law must not conflict with Party priorities or restrict the Party). At base, to have a rule-of-law society, a country must minimally have laws, mechanisms for interpreting and enforcing the laws, equitable application of the laws and popular acceptance of the laws' legitimacy. People must know what the laws are before it is possible to say they accept their legitimacy. The state wants all Chinese citizens to know and obey formal state laws. But there are minority people who haven't yet fully learned and accepted state laws. Many of these groups in fact have their *own* set of laws that developed in their isolated communities over the course of history, and which may well directly contradict state law. The Chinese government worries that imposing state law on minorities who have not yet accepted it or prefer their local minority customary law may have an unintended backlash, and may weaken their willingness to adopt state law. It could even cause minorities to challenge state authority or harm the "social harmony" so important to CCP control. Rule of law demands fair and equal application of the law, and it is certainly problematic to

apply state laws unevenly or have laws on the books that are not applied at all when someone opts to use a contradictory law. Nonetheless, based on my fieldwork, it seems that flexible application of state and customary law may be necessary to promote rule of law in the long run. The state does need to provide citizens and legal workers better guidelines on how this should be done, however, and also to clearly articulate that minority law must conform with broader international human rights norms.

Who are these minorities? Fifty-five different groups are recognized as official ethnic minorities. To be recognized as such, each has to have its own unique territory, language, economy and culture. Historically, the minorities were not at all well integrated into the broader Chinese state. Though certainly some minority citizens have moved to cities, particularly over the last one to two decades, they tend to live in more rural areas, often in very remote mountain or desert regions. Though infrastructure (both physical and communication) is more and more reaching minorities, some communities still have limited interactions with those outside of their minority group and little engagement with non-minority state actors. It is in these types of remote regions that minority customary law tends to hold the most sway, guiding people to act according to local ethnic customs that are binding and can be enforced through a local minority actor such as a village elder or religious shaman. These are minority customary laws, and may be written or just orally conveyed.

My first goal in the field was just to track down examples of conflict of law cases, so that I could then hopefully follow up with interviews of those involved in the cases. This proved *extremely* challenging. Often, I discovered, minority customary law was used, but it was used outside of the formal court system. In those instances, there was no written record of the case. I ended up gathering most of my cases from secondhand sources, either from written Chinese scholarly studies or orally in dozens of interviews I conducted in the field with judges, mediators, village chiefs, village elders, police, educators and peasants. Though I did not attend any formal court hearings, I did watch a local judge handle two separate marital disputes in his office. The cases were handled without appeal to formal state law, though one would have been resolved quite differently had the couple appealed to state law. The second case likely could not have been brought before the court. The couple was married by customary minority law and thus

did not have a marriage license. In the midst of a divorce now, the woman wanted the judge to apply state law in dividing the couple's assets. The judge negotiated an agreement between the two, noting that they couldn't take the case to court since they weren't legally married by the state.

I gathered written materials wherever I could. These included several "village regulations" which are written village laws, often containing articles that conflict with state law. Some, for example, impose heavy fines for petty theft, outside of the parameters allowed by state law. The Chinese government runs extensive "Legal Dissemination Campaigns" and employs tens of thousands of workers to promote legal education. I visited several Legal Dissemination Campaign directors, and reviewed their written materials with them to learn how they encouraged villagers to adopt state law. I also photographed and analyzed legal education billboards and murals in villages. I met with village and township officials who helped mediate land disputes, who shared how they blended state and customary laws when deciding conflicts. Police also shared accounts of how they resolved conflicts appealing to customary law when they knew that the state law would have compelled a very different resolution.

I had an opportunity to travel with police officers to remote villages to help teach ethnic minority members about state laws and provide legal counsel. I attended a middle-school legal education session and met with the Vice Principal for Legal Education. Interviewees shared the challenges of getting people to accept state law, and through those discussions I learned about both state and customary laws. I met with the directors of the Elders Association, who were able to reflect on how the practice of minority customary law had evolved over time in their villages.

I took copious fieldnotes during my time in the field, never knowing exactly what might be important some point later as patterns began to emerge and my thesis took shape. Throughout my fieldwork I sought opportunities to cross-verify my findings. Because one source indicated something, did not necessarily make it so. Sometimes this cross-verification can come from written sources, further interviews or personal observations. Only by approaching the issues from a variety of angles and sources, could I be satisfied with my conclusions.

Lessons to Be Learned

First, and perhaps most importantly of all, I have learned through doing fieldwork the importance of remaining flexible and being willing to adapt my research design to fit new discoveries, opportunities and obstacles. Relatedly, I've learned not to panic when things don't work out exactly as planned. Taking copious notes allowed me to make sense of what I had learned as I added new information that enabled me to interpret my initial unexplained, confusing or contradictory discoveries. I've learned the importance of reading as widely as possible before starting my fieldwork for possible research **models**, but have gained confidence in starting my fieldwork without a fully fleshed-out set of specific questions or a detailed schedule of interviews and appointments.

Fieldwork is as much a craft as a science, and I found that I learned and got better as I gained some experience. I also learned how important it is to choose a topic that can feasibly be researched in the time available, with the access available and at my particular field site. I learned something about my own limits and skillset. I initially wanted to write a book on my fieldwork, but the language barriers of working with nine separate minority groups combined with the challenge of really learning enough about the cultures of these groups (particularly the smaller ones who have largely not been written about in English) to really delve more deeply into the details of their customary law practices simply barred me from completing the larger study. Relatedly, I learned that there's always time for a follow-up fieldtrip, and it's not necessary to have all of the answers before offering valuable insights. I also learned to weigh whether the question I was researching merited the risks and time required to complete the project. Any researcher needs to consider this question, whether it be for a single reflection piece for class, an honors research project or a full book-length study. But this question perhaps becomes even more urgent when considering both large financial costs of travel and field expenses, as well as possible political risks of researching sensitive topics.

Though fieldwork can be challenging, frightening and emotionally

draining, there is little more rewarding.

Interested to Know More about the Study Discussed in this Chapter?

Consult the research publication:

> Kaup, K. P. (forthcoming). "Controlling the Law—Legal Pluralism in China's Southwest Minority Regions," *The China Quarterly.*

Exercises and Discussion Questions:

1. What are some important guidelines for selecting your field site? What guidelines are available for determining how many sites to use?
2. How can you remain flexible and willing to redesign your field strategies without losing the theoretical framework that a solid research design provides?
3. Are there methods that could have been used instead of fieldwork or to supplement the fieldwork for this study?

Recommended Resources:

The Congressional-Executive Commission on China (www.cecc.gov): See particularly the Virtual Academy resources and the 2005 Annual Report, "Special Focus on China's Minorities and Government Implementation of the Regional Ethnic   Autonomy   Law."

Notes

1 Baijie 2017.

2 "Uygur Teacher Gets Life in Prison," 2014; Phillips 2016. https://www.theguardian.com/world/2016/oct/11/ilham-tohti-uighur-china-wins-nobel-martin-ennals-human-rights-award.

3 Read 2010: 148.

4 Manion 2010: 181–199; Gustafsson and Shi 2006: 136–152.

5 Heimer 2006: 72.

References

Baijie, An. 2017. "Xi Advocates Common Security." *China Daily*, June 12, 2017. www.chinadaily.com.cn/world/2017xivisitskazakhstan/2017-06/12/content_29705208.htm

Congressional-Executive Commission on China. 2005. "Special Focus 2005: China's Minorities and Government Implementation of Regional Ethnic Autonomy Law." In *Annual Report*. Washington, DC: US Government Printing Office.

Gustafsson, Björn, and Li Shi. 2006. "Surveys – Three Ways to Obtain Household Income Data." In *Doing Fieldwork in China*, ed. Maria Heimer and Stig Thogersen. Honolulu: University of Hawaii Press: 136–152.

Heimer, Maria. 2006. "Field Sites, Research Design and Type of Findings." In *Doing Fieldwork in China*, ed. Maria Heimer and Stig Thogersen. Honolulu: University of Hawaii Press: 58–80.

Manion, Melanie. 2010. "A Survey of Survey Research on Chinese Politics: What Have We Learned?" In *Contemporary Chinese Politics: New Sources, Methods, and Field Strategies*, ed. Allen Carlson *et al.* New York: Cambridge.

Phillips, Tom. 2016. "Ilham Tohti, Scholar Imprisoned for Life by China, Wins Major Human Rights Prize." *The Guardian*, October 26, 2016. https://www.theguardian.com/world/2016/oct/11/ilham-tohti-uighur-china-wins-nobel-martin-ennals-human-rights-award.

Read, Benjamin. 2010. "More than An Interview, Less Than Sekaka: Studying Subtle and Hidden Politics with Site-Intensive Methods." In *Contemporary Chinese Politics: New Sources, Methods, and Field Strategies*, ed. Allen Carlson *et al.* New York: Cambridge.

"Uygur Teacher Gets Life in Prison." 2014. *China Daily*, September 23, 2014. http://europe.chinadaily.com.cn/china/2014-09/23/content_18646711.htm

Interviewing in Political Science Research

Who Resists Injustice?

Kristina Thalhammer

CONTENTS

Getting Curious: Who Stands up to Dictators?

On May 10, 2017, tens of thousands of women of all ages held high white headscarves as they marched on the streets in Buenos Aires, Argentina, protesting reduced punishments for convicted human rights abusers. These women mirrored and amplified actions taken in this same place for the past 40 years, and their cause in 2017 was the same as it was when the marches began – holding accountable the people who used the government to torture and kill its citizens.

This story begins with the military junta that controlled Argentina from 1976 through most of 1983, authorizing the kidnapping of tens of thousands of regular people, including students, taken from their homes, schools and workplaces and never heard from again. Those suspected of disloyalty were denied their rights to legal processes and never had a chance to prove their innocence. Held in hundreds of clandestine prisons (we later learned), the unofficial prisoners included pregnant women, whose newborns were sometimes placed in the homes of their captors and raised by the very people who had killed their mothers. Most prisoners were brutally tortured. Some were drugged and thrown into the Atlantic from helicopters, their bodies washed ashore in neighboring countries or never found. Virtually everyone in Argentina was afraid to criticize the regime. Yet, in this context emerged human rights activists, who demonstrated against these abuses in broad daylight, right in the heart of Buenos Aires.

Demanding fair treatment for those "disappeared," demonstrators drew domestic and international attention to the abuses and helped apply pressure on the military junta; this eventually led to a dramatic decline in the number of disappearances. Some scholars credit them both with having brought about the international pressure that ended the brutality and with keeping open a small but highly visible public space within Argentina that became crucial in ending the disappearances and helping the country return to democracy.

The nature of the disappearances, the fact that the abuses went on for years, and the generally passive response of most Argentineans to state agents seizing innocent people and refusing to reveal their fates, reminded me of my own family's history and the unanswered questions it inspired. My mother was a child in Hitler's Germany. Her family survived Nazi repression by behaving as most other Christian families did: not speaking out against the regime, even as Jewish, Roma and other purported "state enemies" were rounded up and sent to death

camps. While my grandparents took risks to help one victim, they did not risk taking a public stance against the Nazi policies. Knowing that critics of the Nazis could also be apprehended, should they have done more? Would I have acted differently?

The Argentine regime's brutal treatment of its citizens raised many questions for me. Among them were: How could otherwise moral people allow this to happen? Who on earth *would* persist in facing a ruthless government to try to stop an immoral policy? How were such people different from the rest of us? This final question became the core of my research agenda.

What I read about the Argentine resisters amazed and inspired me. They documented and publicized the abuses, filed *writs of habeas corpus*, and urged international figures (ranging from the Pope to rock stars and even school-children) to pressure the ruling junta into stopping the disappearances and bringing justice to those assumed to be held in clandestine prisons. The most famous of the Argentine human rights groups was the Madres (Mothers) de Plaza de Mayo. This group of middle-aged women, wearing distinctive white headscarves, gained international visibility by holding weekly public marches in Buenos Aires' most popular plaza, right across from the presidential palace. These women, whose adolescent and young adult children had been disappeared, first met as they visited police stations and other government offices looking for answers about their loved ones. After months of fruitless probing, some of them united and decided to demonstrate, trying to draw attention to the disappearances. Others joined them as they came to march every Thursday near the presidential palace. They first asked and then demanded that those who had been disappeared be proven to be alive. The Mothers of the Plaza continued their protests throughout the worst periods of repression, even as members received death threats, were attacked by police, and after a number of the group's founders and several French nuns working with the group were disappeared, tortured and killed. The Mothers and members of other Argentine human rights groups never resorted to violence. How could anyone be brave enough to face such brutality and keep coming back?

My first instinct was to attribute this bravery to "maternal love." I speculated that women can't help but risk everything to protect their children. One of the earliest accounts I read of the Madres (an early draft of Jean Bethke Elshtain's

Antigone's Daughters 1996) also asserted this gender-based explanation. But if that was the case, then why hadn't *all* the mothers of those taken prisoner joined the protesters? As I learned more about what happened in Argentina, I learned that seven other human rights groups also worked within that repressive context, trying to stop the regime from terrorizing its own population. Those groups included many men and women who were *not* related to the missing. They documented disappearances and pushed the courts, politicians, nongovernmental and intergovernmental organizations to apply pressure on behalf of the missing. My early, simple hypothesis that protestors were women motivated and made fearless by maternal instinct or family bonds explained neither why only some affected mothers joined the human rights movement nor why some who had not lost loved ones risked everything to fight the abuses. Members of all the human rights groups were targeted by the regime. Why did so many risk so much for friends, coworkers or even strangers, while many others remained silent?

Eventually the military lost control of the country, brought down by its own incompetence and a powerful protest. Hundreds of thousands joined the Madres in the streets in 1982, demanding a return to democracy. Following the regime's fall and the democratic election of President Raul Alfonsin in 1983, the government took the unprecedented steps of trying the military leaders and creating an official truth commission to document the abuses that occurred under the regime. The military's code of silence limited investigations and the fate of most of those disappeared remained a mystery. Presidential decrees soon freed the convicted junta members, limited further and civil suits against alleged perpetrators, and stated that military below a certain rank were simply following orders and could not be prosecuted. With the abuses unresolved, the eight groups remained intact, continuing to pressure each new presidential administration for information about still-missing loved ones and for policies that would bring appropriate punishment for their torturers and murderers.

I read extensively about the regime and the human rights groups. While the books and articles expanded my understanding of the events, none answered the question of why activists had done what they did, despite the fear-evoking context. I resolved to investigate this puzzle myself and nearly ten years after the regime had collapsed I did so, heading to Argentina to conduct interviews and returning again in 2014 for a follow-up study on the long-term impact of such

activism.

The Research Strategy: Using Interviews to Solve the Puzzle

To explain the courageous resistance of human rights advocates, I didn't just want to analyze existing interviews or to interview activists and summarize their responses. I wanted to know how they differed from the people who did not engage in human rights activism during this period. I needed to locate people who fell into two large categories (i.e., the courageous resisters and non-resisters). Further, to give enough attention to the impact that might come from having a family member seized by the military, I wanted to break apart these two categories. By creating four types, I could compare those whose families had been directly affected by regime violence and those whose families were relatively unscathed, and those who had taken public action with those who had not. To make the comparisons I needed, I decided to interview those who fitted into the following types: Group 1: Those with relatives affected by the violence who took action against the regime (whom I came to call *Advocates* such as Mothers of the Plaza); Group 2: Those with relatives affected by the violence who did not take action (*Beholders*); Group 3: Those with families not directly affected by regime violence who did resist the regime (*Altruists*); and Group 4: Those whose families were not affected by regime violence who did NOT resist the regime (*Bystanders*) (see Table 5.1).

Groups 2 and 4 would serve as comparable groups for Groups 1 and 3 respectively. I could also compare the members of Groups 1 and 3 (all the Courageous Resisters) with the members of Groups 2 and 4 (the Non-resisters). Ideally, I would find similar numbers of subjects with fairly comparable demographic characteristics for each of the four groups and ask them the same questions. Differences in the proportions of the group members answering in particular ways would signal how the groups were similar and how they differed from one another.

In other words, my strategy required me to identify a comparable group that

would differ primarily in the **dependent variable**: the level of nonviolent political activism in this repressive context. Interview questions (including both structured and semi-structured) would help me to gather information about **independent variables** I hypothesized might explain activism. To compare these variables, I would code the responses to various questions and use simple statistical tests of significance (e.g., chi-square) to see which factors differentiated each of the four types from one another (especially from their respective comparison groups) or distinguished courageous resisters from non-resisters.

Table 5.1 Categorization of Respondents on Two Dimensions

		Affected by regime violence	
		Yes	No
Active in human rights movement during peak of repression	Yes (Courageous Resisters)	Advocates (N=27)	Altruists (N=23)
	No (Non-resisters)	Beholders (N=11)	Bystanders (N=17)

Issues with this Research Strategy

One disadvantage of interviewing people about past events is that it is difficult to gauge the impact that subsequent life experiences have had on their memories. Despite my best efforts to help people recall their experiences accurately, I sometimes detected problems during the interviews. For those who were part of the human rights organizations, particularly some members of the Mothers of the Plaza, it was clear that they had told the story of themselves and their organization many times before. Though they had distinct stories about their lives before the coup, a number of them used virtually identical wording to talk about "their" experiences in the early days of the organization. They seemed to share a "script" that included similar words and sequences of activities, always delivered in the first person plural – whether they had actually been at an event

or not! For example, several women reported a narrative like the following: "Then we went to stand in front of the president's house. We didn't know then that he knew what was happening. We thought if we could draw his attention to the fact that our children were missing, he would help us." When pressed on whether each had personally been there for these early demonstrations, most acknowledged that they had joined months later, when the Mothers were already walking around the obelisk in the square and no longer believed that the president was naively unaware of what was happening. These women were not deliberately trying to mislead me. Having told the story of the organization so often, they had formed a collective narrative that dominated their memories, and only with prodding could they again individualize their stories to tell about what each of them specifically had experienced, thought and done. Even when these women did talk about their own experiences, I wondered how much their stories were altered by having shared the common story and experience of being a member of the close-knit core of one branch of the Mothers of the Plaza.

Despite these shortcomings, there is no better way to learn about how people decide to become courageous resisters than by asking them about themselves. Comparing responses of resisters to those of non-resisters and comparing the four groups (using tests of statistical significance) led to interesting results.

The Literature Review

Research trying to explain inhumane treatment expanded dramatically after World War II as scholars sought to understand how people could inflict such unprecedented levels of human suffering on others. Some research looked at the psychology of leaders and followers, some at the structures of situations in which individuals played particular roles or felt themselves subject to others' orders, seemingly absolving them of responsibility.[1] Yet even in studies that demonstrated appallingly high levels of cooperation with unjust authorities by average civilians, significant numbers of individuals defied orders to harm others. Because these courageous resisters were not the focus of these experiments and studies, however, their behavior was not well explained. In addition to trying to

explain why perpetrators of human rights abuses are able to do the unthinkable, some researchers emphasized the important role that passive bystanders play in encouraging oppression and even repression.[2]

The twentieth and twenty-first centuries are full of other examples of resistance to even the most violent acts of state and societal repression, some of which succeeded in saving lives, changing policies and even driving bloody dictatorships from power. Unfortunately, scholars considering these movements rarely look at individual actors and how they became mobile.[3] Little research had been done on why individuals refuse to go along with authorities when they demand compliance with unjust demands such as torture and extralegal executions. The experiments on prosocial behavior and altruism seemed distant from the life-endangering context of Argentina under the military regime. Other than biographies of famous activists and some systematic studies of those who became active in the US civil rights movement, I could find little research explaining why people engaged in high-risk, other-oriented public action. Very few researchers shed light on real-life episodes of other-oriented, high-risk action, or suggested characteristics that might distinguish courageous resisters from non-resisters. Scholars who wrote about the period of Argentine repression similarly failed to address why some people mobilized while others did not, although a number of scholars' research did shed light on human rights groups' histories, strategies, goals or impact.[4]

To generate hypotheses about what might affect individuals' behavior in the Argentine context, I read research on socioeconomic and political mobilization and civic voluntarism, as well as psychology literature. I focused on action in fear-evoking contexts and altruistic behaviors. I was especially intrigued by psychological experiments that explored factors influencing altruistic behaviors.[5] Experimental settings cannot recreate a context of real and ongoing danger comparable to life under authoritarianism, however. Only two studies of altruism I came across compared altruists who had faced real peril to bystanders in that context. Decades after the Holocaust, Samuel and Pearl Oliner conducted interviews with almost 700 people who had lived in Nazi-controlled Europe, most of whom had rescued Jews.[6] While the rescuers' actions were private rather than public, their activities were other-oriented and very high risk, paralleling the Argentine resistance in key ways. The other study, by Kristen Renwick Monroe

(1996), also looked at action during the Holocaust, but compared the life histories of ten European rescuers with those of passive bystanders and others. Perhaps the factors that distinguished rescuers from non-rescuers in that dangerous setting might also explain action and inaction in Argentina.

My survey of the Oliners' and others' research yielded 65 possible factors that might influence resistance or acquiescence to the regime's policies. I reduced these into a handful of categories and a few hypotheses. I hypothesized that human rights activists might differ from non-activists in some or all of the following ways: 1) their contexts or the resources they had access to, 2) internalized values or lessons learned through early socialization or social learning processes throughout their lifetimes, 3) how they saw themselves in relation to others (did they extend their sense of caring and responsibility for others' well-being quite broadly or see their obligations to others as constrained to their families) and 4) whether their families had been subjected to regime violence via a disappearance, arrest or death.

Conducting the Study

Finding Interview Subjects, Gaining Access and Some Surprises

By the time I was able to interview subjects in Argentina in 1992 and 1993, the military was no longer running the government, but the pain of the last military regime was still raw. Less than a decade had passed and, despite the trial of Junta leaders, none of the perpetrators were in jail. The few victims who had survived secret detention centers and the many grieving family members of the thousands still unaccounted for encountered known and suspected perpetrators on the street, in restaurants and other venues. The publication of the Truth Commission's official report, *Nunca Mas* (Never Again), documented nearly 9,000 episodes of disappearance and torture, but left the ultimate fate of most of the disappeared a mystery. In this context, taking out ads or contacting a random sample of Argentines to ask about what they knew about human rights abuses and how they acted at the time would have been both insensitive and doomed to

failure.

I had little trouble finding subjects who had been active in human rights organizations at the time of the repression. But convincing them that being interviewed was worth their while was a challenge. Because all eight of the human rights organizations still existed, I contacted each, first by sending an advance letter on official university stationery, accompanied by a letter of introduction from a scholar who had published research on the regime. I followed up on this letter when I arrived in Buenos Aires, calling each of the organizations and trying to meet someone in the organization. I presented the letters again in person and asked for the individual and organization's help in spreading the word about my quest to find interview subjects. I explained that I was a scholar, interested in doing interviews to compare the varied life experiences and attitudes of those who had been living in Argentina during the period of military rule. I asked their help in spreading the word among current and previous members that I would like to interview those willing to talk about their experiences and gave them information about how to contact me while I was in Argentina. I also promised to protect the interviewees' anonymity and explained that I could not offer any payment for participation.

Being honest with potential subjects by telling them about yourself and your project is an ethical and practical approach to interaction. People deserve to know who is interviewing them and what is likely to happen to the information they'd be sharing. They are also more likely to trust you and be willing to cooperate if you are honest. By stating that I wanted to compare varied life experiences and attitudes of those who lived in Argentina during the period of military rule, I was being truthful, but also deliberately vague. I gave potential subjects an idea of the kind of questions I would be asking, but didn't reveal too much of what I was seeking to understand. I didn't want anyone tailoring their responses to prove (or disprove) my hypotheses.

I asked almost everyone I met for help in spreading the word to other potential interviewees. Sometimes I was able to make an announcement about what I was trying to do at a meeting or activity where current or former human rights group members were present. Sometimes I was given names and phone numbers of people who might be willing to participate. Being able to say that a friend or colleague recommended them to me made those I called more receptive to

listening to my pitch and to participate. Although not all accepted my invitation, many agreed to meet with me. When we met, I always gave them the letters and my contact information. Before we began an interview, I let them know they could skip questions or end the interview at any time. I also asked them to sign a release form, in which they agreed to allow me to summarize their responses or to quote them, so long as I did not identify them. This policy of quoting but not identifying the source is called being "on the record, but without attribution," as opposed to someone saying something "off the record," which means it is not to be quoted at all. I also left them with printed information about my research and asked them to help me locate others who may have responded differently to the repression. No one called me and volunteered to be interviewed, but some of those I interviewed called me and told me that they had a friend willing to speak with me.

Working through existing human rights organizations proved to be an excellent way to find courageous resisters (those in Groups 1 and 3) who were willing to be interviewed; even some no longer in the movement, who were invited through their colleagues, met me for interviews. The first interviewees often asked pointed questions about who was funding my research and whether I would be sharing my findings with the CIA, but when they recommended to others that they speak with me, the fact that someone they knew had also been interviewed seemed to establish my credibility and led to fewer questions.

As noted, I had designed my study to keep the identities of my subjects anonymous. I told each invited interviewee that I would preserve their anonymity. For some of those interviewed, this seemed to reassure them and likely led them to speak more candidly. Others wanted their names and the names of their missing loved ones to be recorded. Even though I was unable to name them, those telling their stories often were eager to name names and to go on record. This was important to them because victims had disappeared into a killing machine that claimed to have no record of them, while the few released prisoners reported being dehumanized by being referred to solely by number or epitaph. Some met me for interviews in the offices of the human rights organizations, others in coffee shops, where I learned how vital it is to take notes. Recorders can miss dialogue when background noise overwhelms recorded voices, and changing batteries should not interrupt a deeply personal exchange.

Occasionally, I was invited to the homes of the activists, a powerful experience that sometimes extended interviews for hours as the interview subjects pulled out documents, photos and scrapbooks. These unexpected revelations taught me a great deal about the context, society and culture in which resisters and non-resisters lived.

I ended each interview by asking about others who might speak with me. I would usually mention that I was particularly interested in talking with those who were like them in many ways but had decided not to be active during this period. These individuals (Groups 2 and 4) were the most difficult to locate – particularly those who had lost loved ones to the regime and never learned their fates. Those who had participated in groups actively working against the regime were typically proud of their activism or at least used to talking about it, even if they were still suffering from the loss and uncertainty. Those who had not been active in the human rights struggles were typically less practiced at talking about this painful period and likely sensitive about revisiting this time and answering questions about their lack of activism. Those in this category who agreed to be interviewed were without exception referred by friends or relatives.

By using this **snowball technique** of finding subjects to interview (i.e., asking those interviewed to identify other possible respondents), I was able to find people with comparable demographic and socioeconomic characteristics. The drawback of this approach is that it could not yield a representative or random sample of the **population**, so I could not with confidence generalize my findings. Yet, the interviews did suggest patterns that both challenged existing theories and generated new theories about sources of activism.

I interviewed people active during the repression from all eight human rights organizations; these individuals were an elite group in that they had information and experiences that were not known by others. Elite interviews helped paint a more complete picture of what the organizations did during the repressive period and how those in the human rights groups interpreted what they were doing. Interviewing elites and supplementing the interviews with other documents allowed me to draw a more complete picture of how group activism began and evolved.

Elite interviews need not involve a representative or random sample. You identify a target group of people who you believe can illuminate various facets of

your research question. Because of the nature of the information we seek, elite interviews do not require a meeting with every insider. When additional interviews expand our knowledge of key events or information, we should keep doing them. When no new perspectives or information is revealed, it's time to stop. An elite interview strategy alone could not answer the question of what made the resisters different, however. I don't think the activists themselves knew exactly what made them different from the friends, neighbors and family members who chose not to act. Only finding comparable people in all four groups and systematically comparing their answers could signal what was distinctive about the courageous resisters.

Designing Questions, Conducting Interviews and Analyzing Findings

In addition to mapping out what the human rights groups did, I wanted to see whether those who took action differed in their individual resources, contexts, values and personality traits and their own experiences with state terror. I modified a number of questions that had been used in related research as well as writing my own questions to create an interview protocol.

I chose not to design a quick survey that could be administered rapidly to many subjects by a crew of interviewers. Without the funding that supported the Oliners' work, I knew that I would not be able to hire and train interviewers. I had to consider what kind of interview to conduct. Distinct types of interviews can serve different purposes.[7] **Unstructured interviews** are like free-wheeling conversations. These can yield an insider's perspective and are sometimes useful when you don't know a lot about a subject (although you should always try to be as knowledgeable as possible, so you don't waste your subjects' time). Elite interviews often take this form. An unstructured interview can provide fresh ideas and insights, but is likely to wander in unexpected ways and typically won't yield reliable data that can be compared across interviews. **Structured interviews** (those with closed-ended responses) work better when we know a lot about a topic and want reliable data that can be easily compared. If you don't know a lot about the subject you are investigating, however, you might

accidentally construct a closed-ended question in a way that leaves out a key response option, or doesn't accurately capture what you think it does, so it would not yield a valid measure of the factor you are trying to explore. **Semistructured interviews** (i.e., asking consistent **open-ended questions** in the same order), however, "can provide detail, depth, and an insider's perspective, while at the same time allowing hypothesis testing and the quantitative analysis of interview responses."[8]

To maximize comparability on some factors, collect data comparable to that in the Oliners' study, and also allow for gaps in my knowledge about how Argentines responded, thought, felt and acted under the repression, my interview questions combined structured and semi-structured interview techniques. This combination was important for several reasons. First, I wanted people to remember their context and thoughts before they chose whether to resist the regime. Recreating the ways that we used to think and feel offers a great challenge to **validity** (i.e., accurately measuring what we're trying to assess). Life experiences can reshape the ways we remember having felt or acted. Memory is more likely to be accurate when it is placed within a broader context of recalling other events from a particular period. Thus, the sequence of questions was important. I first asked about non-threatening early periods in their lives. For example, I asked them to tell me a bit about the makeup of their households as children and young adults and to describe interactions with the members of their households and other role models before the violence. This helped me learn something about who they were before the repression began and what their early socialization, resources and relationships were like. It also should have made their responses somewhat more valid because they were placing their responses in a specific context rather than relying on a habitual way of talking about an era or looking solely through the lens of their current experiences. Because the questions about their early lives were typically not painful or threatening, they also helped build **rapport** between us, which became important as we continued on to questions that tapped potentially more painful memories. Rapport is a certain level of trust and comfort between the interviewer and the interviewee.

The mixed nature of questions was also helpful when the closed-ended questions I asked appeared to miss the mark. For example, one of the closed-ended questions I used attempted to replicate the Oliners' key scale (a series of

questions whose answers collectively can be interpreted as a measure of a personality trait) of extensivity, but was adapted to the Argentine context. One of the Oliners' key findings was that Holocaust rescuers differed from non-rescuers in their degree of extensivity. Rescuers were extensive in that they were significantly more likely to see members of all kinds of social and political groups in their society as "like me" than were their inactive counterparts. Further, rescuers felt a sense of responsibility and care for more members of their society than did non-rescuers. My adapted version of the Oliners' questions about extensivity worked well for many subjects. They thoughtfully described their view of the social and political groups in Argentina by choosing from the options I proffered. But the same closed-ended questions turned out to be highly problematic for others. When asked to consider various groups within the society (e.g., members of the military, of particular political parties and social classes) and rank them from "very much like me" to "nothing like me," a significant number refused to answer these questions. Those who refused to answer the questions rejected the idea of treating any of the groups as monolithic and often lectured me on how within any group one can find good and bad people.

Table 5.2 Responses to Attempts to Have Respondents Categorize Members of Groups as "Like Me" or "Not Like Me"

X²(p<.01)	Advocates	Altruists	Beholders	Bystanders	Total
Objected to or refused to answer question	11 (50%)	3 (18%)	2 (22%)	2 (12%)	18 (28%)
Answered the question without critique	11 (50%)	14 (82%)	7 (78%)	15 (88%)	47 (72%)
Total	22 (100%)	17 (100%)	9 (100%)	17 (100%)	65 (100%)

At the time this frustrated me. Why wouldn't they just answer the questions as I asked them and move on to the next? I diligently jotted down their critiques, just as I wrote down the answers to the other questions I had asked. Only later did I appreciate that the refusal to accept my categorizing their countrymen was actually more telling than if they had just complied with my original request. This was a good reminder that qualitative interviews can be theory generating. Half of the Advocates (Group 1) I interviewed refused to answer the question as asked. Over three-quarters of Altruists, Beholders and Bystanders (Groups 2, 3 and 4) answered these questions without critique (see Table 5.2). Further

examination showed a dimension I had not run across in any of the research: More of the Advocates, who generally had not been as politically active before they lost family members to the regime, tended not to see groups as monolithic or insurmountable units. At least half saw the organizations and political parties as comprised of individuals, some of whom were potentially sympathetic to requests for helping regime victims. Most surprising were mothers, who knew their children had been kidnapped by the military, but refused to categorize members of the military as "not like me." They argued that even members of the groups that had caused them the most heartbreak deserved to be treated as differentiated individuals. Perhaps this perspective might be a distinguishing characteristic of other courageous resisters. The Oliners found that altruists differed from passive bystanders, not in their access to wealth, space or other resources, but rather in that they tended to have a more "extensive" way of seeing the world. In a similar but much smaller study, Kristen Renwick Monroe concluded that rescuers of Jews were distinct in seeing themselves as one with victims they assisted. My research echoed some of the Oliners' findings. Like European rescuers, Argentinean resisters were not distinguishable from bystanders in terms of their demographic characteristics, physical resources, socialization or social learning experiences. My findings differ from other research in not finding patterns that distinguish those who take action from those who don't. Unlike the findings of both the Oliners' and Monroe, Argentines who acted on behalf of others did not demonstrate significantly different levels of extensivity as measured with the Oliners' scale or in content analysis of their language use. However, it was hard to make this comparison statistically, because about half the Advocates rejected the very idea of categorizing any group in the society as "not like me" and refused to answer the question. The high rate of refusal *did* distinguish the Advocates from all the other groups and suggests they had a distinctive way of looking at the world relative to the others, including other courageous resisters.

The blow of losing a family member did not immobilize all Argentines. It tended to paralyze those who were fairly well informed about the regime's brutality, had more experience both with politics and with fear-evoking events, and were more aware of the risks associated with public activism. These Beholders encountered repression armed with a combination of knowledge and experience, but the disappearance of a loved one apparently made them reluctant

to endure (or subject their families to) more punishment by becoming resisters. In contrast, Advocates also lost family members, but they tended to know less about the level of repression and had less political experience and fewer scars from previous fear-evoking encounters with the state. They tended to respond to their loved ones' disappearances with less fear and more optimism that they could find them (including hope of finding help within organizations others saw as monolithic). Naively going forward, they started searching for one person and wound up starting a movement that sought justice for 30,000.

The naïve Advocates' path to activism contrasts with that of the Altruists, who tended to be experienced political actors. The Altruists typically had survived previous fear-evoking experiences and seemed to be tougher for it. They often described their resistance as a continuation of previous efforts, and went into action against the regime with their eyes wide open to what was happening. Aware of the dangers of activism, they felt they could strategically position themselves and their families to survive dangers even as they remained active. These characteristics distinguished them not only from the Advocates, but from the Bystanders, who were slightly less knowledgeable, experienced or used to high-risk settings.

The impact of recent regime violence against one's family clearly affected individuals' perceptions and responses to the regime. It also interacted with their previous experiences with risk and politics in distinct ways. Advocates, who were more likely to be female, depended on narrow circles of friends and electronic media (manipulated by the regime) and were political neophytes. They likely never would have become political activists if not for the loss of a loved one. Altruists were more buffered from the direct effects of regime violence when deciding to continue their activism. They were more likely to be male, well connected, well educated and used to juggling multiple responsibilities; their political participation was a continuation of public activism begun before the repression. Nearly all (97%) of Altruists described themselves as politically active before the coup, compared with only 30 percent of Advocates; this is significantly different from the scores of Bystanders (47%) and Beholders (64%).

A key finding is that the likelihood of people taking action against regime violence relates both to their previous experiences, perspectives and levels of information and to their relation to the current victims. The interaction between

these factors plays out in different ways. While some of those previously active were frightened into silence by close contact with regime violence, others were inspired to undertake activism for the first time because of that violence (see Table 5.3). Among those not directly affected by current regime violence, some chose to resist the regime in part because they knew the risks, had survived danger before, and were confident that they could handle the new context as they continued their political activism. Those not affected by the regime, who had not been inoculated against fear in this way, often chose the more risk-averse pathway of remaining passive bystanders.

Table 5.3 Recalled Experiences Before 1976 Coup That Evoked Fear

X^2 (p<.05)	Advocates	Altruists	Beholders	Bystanders	Total
Recalled specific episode that frightened them	10 (45.5%)	20 (87%)	6 (67%)	11 (65%)	47 (66%)
No recollection of fearful episode	12 (54.5%)	3 (13%)	3 (33%)	6 (35%)	24 (34%)
Total	22 (100%)	23 (100%)	9 (100%)	17 (100%)	71 (100%)

While I published this first study in 2001, I continued to follow others' research about the struggle for justice there and regularly read journalistic accounts of the ongoing struggles. I was frustrated by the military's continued code of silence and the government policies that prevented prosecution of torturers, kidnappers and murderers. The 2003 election of Nestor Kirchner as Argentina's president signaled the first time since the mid-1980s that the government made human rights a national priority. President Kirchner, Congress and the Supreme Court toppled decades of state-manufactured obstacles to justice. At last the state could prosecute, convict and punish hundreds of members of the military for crimes against humanity. The support for human rights continued as Cristina Fernandez de Kirchner succeeded her husband in office until 2015. As I read reports of the government honoring human rights groups and seeming to implement many of their suggestions (e.g., implementing policies, constructing memorials, circulating curriculum and promoting other mechanisms to rectify the past), I wondered how those who had struggled so hard saw the changes and their future relations with the state.

Thus, in 2014 I returned to Argentina with anthropologist Claire Branigan to see the impact of long-term rights advocacy. Other scholars have written about the legacy of human rights advocacy, but these works examined the movement's

impact in opposition to the state.[9] We were curious how human rights advocacy changed context, networks and people before *and* during the period of cooperation between government and justice-seekers.

We believe that no one would have more intimate knowledge of what human rights groups had sought to accomplish and what they did to achieve it than those engaged in 40 years of sustained activism. We used elite interviews to examine perceived changes in the realms of 1) context (including newly created and modified legal and state institutions, cultural changes and modification of the physical landscape), 2) relationships between activists, society and the Argentine state, and 3) socialization factors that affect attitudes and behaviors of Argentines (including changes in the ways that the era is referenced, taught and memorialized). Juanita Pargament, one of the first individuals I had interviewed in the 1990s, was still a spokesperson for the Association of Madres de la Plaza; she granted us an interview in 2014 as a sharp and articulate 100-year-old! While most of those interviewed expressed appreciation for advances during the Kirchner era, responses were often lengthy, free-ranging and surprisingly critical of certain aspects of contemporary current context. Their responses revealed and confirmed enduring and important legacies of activism, and some confidence that these changes should make recurrence of human rights atrocities in Argentina less likely.

This government, sympathetic with human rights, was replaced with the election of a new president in 2016. In May 2017 the Argentine Supreme Court, influenced by the new president, tried to significantly shorten prison sentences of convicted human rights abusers. Hundreds of thousands of citizens flooded the Plaza de Mayo, many waving the Madres' symbolic white scarf and using familiar chants of "not one step back," reinforcing a rapidly passed law to annul the Court's decision. While this seems to reinforce our conclusion that attitudes and behaviors of the overall citizenry were profoundly changed by the human rights movement, the new context and mobilization raises new questions as well. And so ideas for research continue.

Lessons to Be Learned

Using interviews can help us understand more about the backgrounds, resources and perceptions of the people who behave both in exemplary and more mundane ways. In my research I learned to construct questions carefully and be open to critiques by those being interviewed. Closed-ended questions can speed up the pace of interviews, allowing you to conduct more in a shorter amount of time and with fewer resources. They can also be easier to code and analyze. However, such questions do not always capture what your subjects are really thinking. Semistructured interviews allow those being interviewed to express themselves more freely and often help add a much deeper level of context and understanding to what you are trying to study. They may, however, be more difficult to compare, require more time to code, or defy quantitative summaries altogether. Semistructured interviews can also be difficult to control in terms of time (one of mine stretched to more than 12 hours over several days).

Second, it is not always possible to get a random or representative sample of interviewees. Even if random or representative sampling won't work in a particular context, be as systematic as possible in finding subjects, but be prepared to modify your approach if the context demands. In some cases, especially when trying to speak with experts or a particularly wary population, you may try using the snowball technique to find potential subjects and win their confidence. Be aware, though, that when you abandon sampling techniques, you also compromise your ability to generalize your findings. Depending on your research question, however, elite interviews may yield the answers you seek, with no need for comparison. For example, when I returned to Argentina in 2014 to do research on how human rights advocacy was functioning in relation to a supportive government, we used elite interviews. Those consistently immersed in the decades-long struggle were uniquely situated to describe and assess how they saw the movement's changing relationship with the state and larger society. Third, present those you are hoping to interview with as much information about you, your study and how you will use the material you gain from your interview, then live up to the agreements you make with the people you interview. Fourth, put the needs of your potential subjects first and find ways to let them remain in control during the

interview; allow them to refuse to answer questions or modify or critique questions if they insist. Sometimes people's unwillingness to answer a particular question may mean that you have a confusing or ambiguous question.

My research helped me discover that direct experience with fear and loss can manifest itself both in mobilizing and immobilizing potential courageous resisters. It also brought me into close contact with people who were still suffering from deep losses and led me to question the ethical trade-off of interviewing in such a context. In order to increase our understanding of what happened and share an important example of successful resistance with a broader audience, we need to conduct research. But is it fair to those we interview to ask them to remember periods that may be painful to them to make this happen?

Interested to Know More about the Studies Discussed in this Chapter?

Consult the publications:

Thalhammer, Kristina. 2001. "I'll Take the High Road." *Political Psychology* 22(3): 493–520.

Thalhammer, Kristina, and Claire Branigan. 2017. "Fighting State Terror and Becoming the State." *Revista de Paz y Conflictos.* http://revistaseug.ugr.es/index.php/revpaz/article/view/4944/5591. (This bilingual publication does have the article in English.)

Thalhammer, Kristina, Paula O'Loughlin, Sam McFarland, Myron Glazer, Penina Glazer, and Sharon Shepela. 2007. *Courageous Resistance: The Power of Ordinary People.* New York: Palgrave Macmillan.

Exercises and Discussion Questions:

1. What kinds of research questions do you think can best be answered by using the interview method? What questions do you have that you would like to explore through interviews? Would you use closed-ended or open-ended questions and why?
2. When might interviewing a random sample of the population be preferable to interviewing elites? When might elite interviews be a better strategy?
3. What are some of the challenges to validity that can arise in studying events that happened some time ago? What techniques might increase the validity of responses?
4. How much information should a researcher give about the study they are doing? Is it ever ethical to mislead interview subjects about your research? Is it ethical to revisit painful periods in someone's history for the sake of a study? How can someone who has not lived in a repressive setting study courageous resistance or other exemplary behavior without appearing to criticize the people who quite rationally chose the safer route, such as not challenging a brutal regime?

Recommended Resources:

Aberbach, Joel D., James D. Chesney, and Bert A. Rockman. 1975. "Exploring Elite Political Attitudes: Some Methodological Lessons." *Political Methodology* 2(1): 1–27.

Berry, Jeffrey M. 2002. "Validity and Reliability Issues in Elite Interviewing." *PS: Political Science and Politics* 35(4): 679–682.

Goldstein, Kenneth. 2002. "Getting in the Door: Sampling and Completing Elite Interviews." *PS: Political Science and Politics* 35(4): 669–672.

Leech, Beth L. 2002. "Interview Methods in Political Science." *PS: Political Science and Politics* 35(4): 663–664.

Maynes, Mary Jo, Jennifer L. Pierce, and Barbara Laslett. 2008. *Telling*

Stories. Ithaca, NY: Cornell University Press.

Mosley, Layna. 2013. *Interview Research in Political Science.* Ithaca, NY: Cornell University Press.

Woliver, Laura R. 2002. "Ethical Dilemmas in Personal Interviewing." *PS: Political Science and Politics* 35(4): 677–678.

Notes

1 For example, Zimbardo 1971; Milgram 1974; Kelman and Hamilton 1989.

2 Staub 1989; Goldhagen 1996.

3 For example, Keck and Sikkink 1998; Ackerman and Duvall 2000.

4 For example, Brysk 1990, 1994; Feijoo 1989; Fisher 1989; Guest 1990; Mignone 1991.

5 For example, Latane and Darley 1970; Batson 1973.

6 Oliner and Oliner 1988.

7 Leech 2002.

8 Leech 2002: 665.

9 For example, Sikkink 2008; Jelin 2003.

References

Ackerman, Peter and Duvall, Jack. 2000. *A Force More Powerful.* New York: St. Martin's Press.

Batson, D. 1973. "From Jerusalem to Jericho." *Journal of Personality and Social Psychology* 27(1): 100–108.

Brysk, Alison. 1990. "The Political Impact of Argentina's Human Rights Movement." Ph.D. dissertation, Stanford University.

Brysk, Alison. 1994. *The Political Impact of Argentina's Human Rights Movement: Social Movements, Transition and Democratization.* Palo Alto: Stanford University Press.

Elshtain, Jean Bethke. 1996. "The Mothers of the Disappeared." In *Finding a New Feminism,* ed. Pamela Grande Jensen. New York: Roman and Littlefield, 129–148.

Feijoo, Maria Del C. 1989. "The Challenge of Constructing Civilian Peace." In *The Women's Movement in Latin America*, ed. J. Jaquette. Wincester, MA: Unwin Hyman.

Fisher, Jo. 1989. *Mothers of the Disappeared.* Boston: South End Press.

Goldhagen, Daniel. 1996. *Hitler's Willing Executioners.* New York: Alfred A. Knopf.

Guest, Ian. 1990. *Behind the Disappearances.* Philadelphia: University of Pennsylvania Press.

Jelin, Elizabeth. 2003. *State Repression and Memory.* Minneapolis: University of Minnesota Press.

Keck, Margaret and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics.* Ithaca, NY: Cornell University Press.

Kelman, Herbert, and V. Lee Hamilton. 1989. *Crimes of Obedience.* New Haven: Yale University Press.

Latane, Bebe, and J.M. Darley. 1970. *The Unresponsive Bystander.* New York: Appleton Century Crofts.

Leech, Beth. 2002. "Asking Questions: Techniques for Semistructured Interviews." *PS: Political Science and Politics*, 35(4): 665–668.

Mignone, Emilio. 1991. *Derechos Humanos y Sociedad: El Caso Argentine.* Buenos Aires: Centro de Estudios Legales y Sociales.

Milgram, Stanley. 1974. *Obedience to Authority.* New York: Harper and Row.

Monroe, Kristen R. 1996. *The Heart of Altruism.* Princeton: Princeton University Press.

Oliner, Samuel, and Pearl Oliner. 1988. *The Altruistic Personality.* New York: Free Press.

Sikkink, Kathryn. 2008. "From Pariah State to Global Protagonist: Argentina and the Struggle for International Human Rights." *Latin American Politics and Society* 50(1): 1–29.

Staub, Ervin. 1989. *The Roots of Evil.* Oxford. UK: Oxford University Press.

Zimbardo, Philip G. 1971. "The Power and Pathology of Imprisonment." *Congressional Record* (Serial No. 15, 10–25). Hearings before Subcommittee

No. 3, of the Committee on the Judiciary, House of Representatives, Ninety-Second Congress, *First Session on Corrections, Part II, Prisons, Prison Reform and Prisoner's Rights: California.* Washington, DC: U.S. Government Printing Office.

CHAPTER 6

Critical and Interpretive Research

Understanding Torture's Popularity in the United States

Brent J. Steele

CONTENTS

Getting Curious: How do we Interpret Public Opinion on Torture?

My two primary research fields within the political science sub-field of International Relations are international security (the ways in which states and other international actors 'secure' their communities), and international ethics (the role of ethical considerations in influencing the behaviors of states and other international actors). The topic of torture is one that resides at the intersection between those two research fields, and debates over torture were reignited when I was in graduate school, after the 9/11 attacks and the subsequent war-on-terror policies that were instituted in the early–mid 2000s. Torture is often defended as being necessary for the security of political communities, as important information that can be extracted from a tortured subject can then be used to save lives. When I first started researching torture in the mid-2000s (when more documents regarding the US's practice of it started becoming available), most of the arguments against its use were 'ethical' or 'legal' arguments: that it was immoral to torture individuals, or that it was illegal in international and US domestic law. I teach an international ethics course and most of the time the 'pro-torture' arguments fold into the former category, and the 'anti-torture' ones into the latter. However, as I have researched torture over the years, I've found quite a number of sustained arguments against not only the ethics (it's morally wrong), but the *efficacy* (it just doesn't work), of torture as well. Nevertheless, torture continues to be supported by majorities of US Americans in survey after survey. This led me to publish a series of studies investigating what factors may be responsible for its increasing popularity, including a chapter in an edited volume that specifically focused on the trend emerging in **surveys**, the results of which I focus on here in this chapter as well.[1] While I oppose torture on both moral and strategic policy grounds, I am also quite interested in why it continues to be popular amongst US Americans.

A conclusion I've drawn in those studies is that one of the biggest problems with torture and its use in twenty-first-century global politics, especially within liberal democracies (like the United States), is its *hiddenness.* One of the most infamous accounts of torture and punishment comes from poststructural theorist Michel Foucault.[2] Foucault argues that "punishment as a public spectacle," including especially torture, declined in the nineteenth century in many Western countries and was transformed into the "most hidden part" of the legal process. Yet it is not only torture itself, when it is practiced, that remains hidden from our

analyses. Surveys regarding torture give us an indication whether respondents support its use, but do not necessarily tell us why (or why not) they might support it. Nevertheless, torture is becoming more popular in the United States. And the further in time one gets from 9/11, as evidenced by surveys taken at the end of the 2000s through the mid-2010s (see below), the *more* popular torture has become. Contrary to what we would expect in more 'open' societies where policies can be debated and, presumably, the more rational policies will win out, the *more* the topic is debated, the more popular torture has become. All of this has occurred in a liberal democracy supposedly known for 'upholding human rights,' due process, civil liberties, *etc.* Furthermore, such popularity runs counter to the predictions (and even observations) of notable experts on torture and democracy that I discuss in this chapter.

The Research Strategy: Critical and Interpretive Methods

How can critical and **interpretive methods** help us on this topic? Surveys definitely provide us a good start, but they have only limited utility if we don't use them in conjunction with other approaches and methods. Critical and interpretive research strategies more broadly are intended to call scholars' (and citizens') attention to voices, processes, people, groups, materials and ideas that other methods and perspectives often overlook or, even more perniciously, exclude or selectively ignore. Critical and interpretive research strategies are useful not only for *what* they are calling to be studied, but also *where* one can study it.

Following William Connolly, I refer to these critical and interpretive methods as "**micropolitical**" in their approach.[3] They are a group of research strategies that approach smaller settings because they assume that these contexts are where political views get formed and reinforced, including churches, schools, cafes, town hall meetings, sporting events, as well as through films or other forms of popular culture. They are, in other words, a collection of methods – from ethnographic research to film analysis to focus groups – that are useful for

delineating, and then analyzing the narratives that people use to make sense of the world. Thus, the primary interpretive research method of *narrative analysis* is itself a product of these micropolitical methods.

Such methods can, I argue in this chapter, be particularly useful as a complementary set of methods that help scholars think about the ways in which quantitative methods fall short in explaining. In being 'critical,' such methods ask us to critique not just for the sake of critiquing, but rather to dig more deeply into the processes that our conventional methods are used to understand, and question whether those reveal, as the iconic US radio host Paul Harvey used to say, "the rest of the story." In being 'interpretive,' such critical methodologists argue that interpretation is a part of the world of politics, and that understanding how those interpretations not only come to be, but come to dominate, is key to understanding international politics itself.

Regarding the issue of torture in particular, surveys help to reveal just enough about the trends in the United States to show that torture is getting popular, but they also are limited in that they don't tell us why this is the case. Being 'micropolitical' in one's approach, I argue, can lead us to locations and sites where the 'logic' of torture gets fostered, facilitated and reinforced.

A brief word about the format of this chapter since, considering the uniqueness of the methodological discussion, it deviates slightly from the format of the other chapters in this textbook. As the case for micropolitical research very much rests on the insufficiency of conventional survey methods, the chapter will first review the literature on public opinion and torture in the US ("Literature Review: Survey Says …"). This will include a discussion on how the aggregation of such data actually *conceals* (a form of hiding) the popularity of torture in the US through time. I then discuss some theoretical and methodological strategies for examining the narratives of torture ("Doing the Study: (Critical) Micropolitical Analyses as a Complement to Surveys"). In the last section ("Lessons to be Learned"), I will discuss some shortcomings of micropolitical methods, and this involves the disclosure of some of the political purposes of critical methods more generally.

The Literature Review: Survey Says (or not?)

Why are micropolitical methods necessary on this topic? Conventional research strategies utilizing surveys have their own limitations. The surveys and data selection on state-practiced torture, from the surveys of respondents – which remain rather infrequent – to the information about torture policies as they have been practiced (most of which remains classified and, in some cases, destroyed), remain limited and infrequent. Further, surveys are usually carried out when the topic of torture is 'hot,' on the front burner of news sites and in social media, but rarely conducted during times when, as a topic, it fades into the background of other news. Thus, the topic of torture is itself a limitation for those scholars working to understand its broader appeal, at least in the US. Further, because it is conducted 'in the dark,' important documents or data that might reveal how it was carried out and what it produced remain classified or redacted, and thus inaccessible for the researcher. Claims by authorities that 'torture works' or has worked in the past, are difficult to scrutinize, let alone verify or refute, and thus study. The broader problems with government documents and source access are especially relevant on this topic. In the following I shall discuss these issues as they appear in the literature.

There are a number of studies on public opinion and torture, and digging into these and how they are presented to the public as well as how they are treated by experts and scholars proves important. Both Darius Rejali and Paul Gronke, in a series of articles or essays, have made the claim that torture is, overall, *not* popular in the United States.[4] In one essay, they remark, unequivocally, that:

> The American public has spoken in 30 polls since September 11, 2001 that it does not believe that the use of torture is justified, even if it works. What is reflected in the polling data is that the majority of Americans support the principles of fairness and decency, even when there are more expedient means at our disposal.

In a more recent article, they and several other co-authors claim that:

> This stance was true even when respondents were asked about an imminent terrorist attack, even when enhanced interrogation techniques were not called torture, and even when Americans were assured that torture would work to get crucial information.[5]

These assertions are no longer accurate, and even at the time (in 2010) they were based on a selective, static, categorization of "the [US] public's attitudes on torture." Rejali and Gronke get to this conclusion based on aggregating *all*

surveys from 2001 to 2009, rather than observing how the responses changed over that period of time. Even more recent studies have done this – most likely for the reasonable purpose of increasing sample size for 'statistical' analysis – and yet they all repeat the claim, without caveats, that "more Americans [are] opposed than supportive" of torture.[6]

Aggregation of surveys over a decade has its strengths. As noted earlier in this chapter, it is difficult to generalize about torture's popularity since there are a smaller number of surveys regarding it compared to, say, approval of presidents. But aggregation also misses some of the trends clearly visible in recent surveys. As the surveys that Rejali and Gronke aggregate in their own Table 1 demonstrate, torture has become more and more popular over time in the United States.[7] A survey of over 1,000 US respondents conducted by the ORC International for the American Red Cross in March 2011 found that over 51 percent found torture to be "acceptable at least sometimes,"[8] and a set of Pew surveys demonstrated a strong majority (54–41) of respondents saying that torture is 'often' or 'sometimes' justified against 'suspected terrorists' as opposed to those who answer 'never' or 'rarely.'[9]

In my interpretive chapter, I argued that one should be careful of the manner in which these responses are collapsed – in that doing so in this way almost *conceals* torture's popularity. Here again we see how there are problems with the hiddenness of torture; in this case, how we aggregate survey responses causes us to overlook its increasing popularity. Since the United States, in ratifying the Convention Against Torture in 1994 (it was signed in 1988), made torture illegal under *US law*,[10] one might note that even respondents who say that torture is 'rarely' justified are in fact sanctioning an illegal act, not to mention that if agents in power are given the option to torture 'rarely,' they can do so without public sanction consistently. When we conflate the 'often,' 'sometimes,' or 'rarely' response options offered in most surveys regarding when torture is 'justified,' the number in support of some kind of torture jumps to 70 percent – the highest level it has ever been. I might note here that the phrasing of the questions is *not* at issue – all these surveys use the term 'torture' rather than the popular 2000s' euphemism of 'enhanced interrogation techniques.' Respondents are expressing their support for *torture* in these surveys, not techniques 'short of' it.

Further, and in a theme I will return to below, my interpretive analysis from my research chapter asked if scholars have failed to consider fully the role of religion in shaping opinions of torture. In terms of those who state that torture can be 'often' or 'sometimes' justified (with 'rarely' added in parentheses) we have the following figures, from a 2011 Pew poll:

- White Evangelical Protestants – 62 percent (79%);
- White non-Hispanic Catholics – 51 percent (73%);
- White Mainline Protestants – 46 percent (68%).

Yet the more important and statistically viable relationship which seems to come through in this same study is that those who attend religious services most frequently are *most* supportive of torture – 54 percent (73%).[11] 'Unaffiliated' individuals, and those who attend religious services 'seldom' or 'never' are *least* supportive of torture – 42 percent (69%).

Two caveats should be noted here that were discussed in the earlier 'limitations' section of this chapter. First, the connection between religiosity, religion and support for torture is not completely clear cut. The Pew poll's sample size overall, and its sub-sample categories broken down by religion, are rather small.[12] Other studies, further, find that the connection between religion and support for torture is 'mediated' by 'political alignment,' which suggests that in the US polity especially, this is what happens when both torture and religion are politicized in a highly partisan, and even tribal, democratic society.[13]

Gronke *et al.* do acknowledge the possibility of torture's increased popularity, but they also respond that this increase is due simply to partisan politics:

> The appearance of a public majority who favors torture is a very recent phenomenon. We believe that torture may have become a partisan symbol, distinguishing Republicans from Democrats, that demonstrates hawkishness on national security in the same way that being supportive of the death penalty indicates that a person is tough on crime.[14]

Indeed, as one Republican 'blog' asserted, rather triumphantly, after the November 2009 Pew results were released, "In case you're wondering who won the de facto debate between Obama and Cheney, wonder no more." Yet in that same post, it rightly points out the following wrinkle for the 'partisan politics' hypothesis, also found in the Pew poll in November 2009: "Republicans are +2 [in

supporting torture] since Obama was sworn in, independents are +9, and Democrats are *+18*."[15] Something is at work here that goes beyond 'symbolic partisan' politics.

I made the argument in my 2013 study that the political assumptions of scholars may have played a role in overlooking this increase in public support for torture. What I meant was that, as political scientists working (relatively) freely within an open democratic society in the US, some may have a base view of a 'rational public' that they assume is far more astute about torture than the elites who lead it. Paul Gronke and Darius Rejali's interpretation (going further than just a survey observation), for instance, was that "the majority of Americans support the principles of *fairness and decency*, even when there are more expedient means at our disposal." Surveys rarely, if ever, ask questions about 'fairness and decency,' but scholars like Gronke and Rejali intuit that this is what the public supports on the issue of torture. It is *here* where micropolitical methods can help access just how 'fair and decent,' just how influenced (or not), a public is by the narratives used by elites to justify torture.

Again, owing to the critical normative purposes noted earlier in this chapter, the broader debate here is not only methodological and empirical (what do the surveys show?), but theoretical and normative-ethical as well. Theoretically, there is an assumption – held at the core of liberal democratic theories – that the 'public' is right, that it will over time balance against the excesses of elites or policy, and that it is rational. Support for torture will fall once the public 'sees the light' and can utilize that sight to formulate its voting behavior and democratic participation.

But this hasn't happened. As I discuss in the conclusion to my most recent work on this method, the 2014 US Senate Intelligence Report debunked every torture claim or myth circulating at the time about the efficacy of the CIA's torture practices of the 2000s. And yet in polls taken following the report's publication, the American public agreed in two polls by almost 2–1 margins that the torture of suspects by the CIA was 'justified' and produced 'valuable intelligence.'[16], [17]

Because these 'surveys' show us something different – that the public is neither passive nor proactive but can be *activated* and then participate in supporting these policies – then the public must share some of the *responsibility*,

some of the agency, for the recent practice of torture – and its fallout and outcomes. But how can we understand what makes this support possible? Instead of looking at these surveys as a 'why' question, I suggest a more modest approach: to look at them as a 'how possible' question, and look at what possible practices or tendencies that make support possible, but not determined, for these particular segments of the US polity.[18] What is going on here is the 'power of counterfactual reasoning,' of epistemic 'faith' – which is to say that Americans do not *know* whether torture really produces truth, but they believe that it does. To locate that reasoning is a challenge, but one that critical micropolitical methods may just be up for.

Conducting the Study: (Critical) Micropolitical Analyses as a Complement to Surveys

While surveys help us ascertain the direction of torture's popularity in the United States, they do not give us a good idea why, or just as importantly, *where* its popularity gets formulated and justified. The move to micropolitics brings more voices into the conversation on torture. Micropolitics refers to "those features of social life that often slip through our normal schematic or binary frameworks. Sensation, resonance, movement, flow … features that often form the unspoken but experiential constitution of our larger categories of nation, state, economy, security and so on."[19] Micropolitics is not its own method, but a category of methods or approaches to extract the narratives that we want to analyze to discern why and how, in this case, torture's popularity in the United States has increased. Participant observations, ethnographic methods, and focus groups are all examples of methods that might be used to ascertain those phenomena that are missed by broader **quantitative analysis** or methods.

One example of a micropolitical approach is Katherine Cramer Walsh's study that observed, via the method of focus groups, rural Wisconites to assess their views of politics.[20] As Walsh succinctly concludes, "[T]he study suggests that public opinion research more seriously include listening to the public."[21] More directly related to security issues, the research on "vernacular security and non-

elite knowledge" by Daniel Stevens and Nick Vaughan-Williams provides another methodological avenue for analyzing "everyday" narratives on torture.[22] Stevens and Vaughan-Williams use "juxtaposition," which entails "comparing and contrasting what different participants [interviewees] said (and did not say); and by setting these diverse opinions against the backdrop of dominant and otherwise homogenous (and elitist) national security frames."[23]

In the context of torture, micropolitical research strategies can investigate and channel the voices of those who have been excluded or (again) *hidden* from public and often elitist discourses. This includes those who know someone who has been tortured, or have been tortured themselves. Such sensitive but important work has been done for years by organizations like the Center for Victims of Torture, a St. Paul-based global NGO that not only collects narratives of these groups but also provides a checklist for doing ethnographic work with victims of torture.[24] Yet we might also locate the voice of the 'everyday' citizen who supports torture and discusses it in a variety of settings (like churches, classrooms, political rallies, coffee shops or pubs). This type of research begins with a variety of methods, including ethnography, focus groups, narrative analysis, and perhaps (ethical, controlled) experiments. Such methods, then, can be used, not only to trace the narratives that get invoked to justify torture and find out their origins, but also to discover the modifications that subjects use to 'make sense' of torture.

The types of methods needed to explore this thread are more interpretive, that is, they do not easily lead themselves to just 'one' reading or conclusion but usually have to be teased out in all their complexity. Of course, all methods require, at some point, interpretation. But these methods can be less 'subjective' and more focused and situated if they are supplemented and shaped by particular theoretical arguments. These arguments help reduce the otherwise (too) open field of topics that might arise in such settings and bring some order, but not too much rigidity or formality, to our investigations.

Three potential mechanisms may be at work here in increasing torture's popularity, and discussing them ahead of time might help sharpen our focus for the micropolitical methods that follow. The first of these we might title '**narrative extrapolation**' – a discourse that helps to specifically narrate a causal story. Put another way, such extrapolation comes from 'causal narratives' that

can serve to 'teach' individuals the logic regarding particular policies, including torture. What is the theoretical basis for focusing on narratives? The importance here is that politics is social – it is the stories that we tell, and that are told to us, that help shape our opinions and perspectives. Further, it is where we tell them (as noted earlier) – at that coffee table conversation with members of our community, at a reunion with family friends and relatives, in the church basement after hearing a sermon or homily, or at a bar or bowling alley, and how we internalize a script of these stories as if the story was *our own.* As Hannah Arendt asserted, we realize our own agency – our own participation in the public realm – through the ability to narrate a story: "in acting and speaking, [people] show who they are, reveal actively their unique personal identities and thus make their appearance in the human world."[25] Torture 'stories,' like all political ones, become less hypothetical and more (inter)personal for subjects discussing their support.

A second, related mechanism that is itself a value, that perhaps makes torture popular for certain respondents, is *faith*, and its relationship to 'intent' – a faith or trust in authority by those who claim a positive/beneficial 'intent' for their torture. Intentions are often central to the justification of US-style torture. For instance, if the intent behind such torture is to 'save lives,' it is justifiable. In this vein, we might find how one US Christian evangelical politician – Gary Bauer – in an essay subtitled "How Christians can think about the unthinkable," defends torture:

> [t]he issue which has been ignored to date in the discussion of enhanced interrogation is whether there is a difference between inflicting pain for its own sake or using some harsher methods with deliberation when lives are on the line. When the *intent* is to extract information necessary to save human beings in imminent danger, harsh treatment may be justified and, I believe, sometimes necessary.[26]

There are two aspects of a focus on *intent* that need to be extracted here in order to understand both the persuasiveness of the torture fantasy and also its fluidity and flexibility, and therefore to suggest some investigative avenues when doing participant observations. The first is that intentions are (again, like much of torture itself) *hidden* from our view – they can be politically (and after the fact) manipulated quite easily into various purposes or ends. Kept hidden from our view, the intent behind torture itself can therefore be reconfigured – especially if we have a 'faith' in that which we cannot see (intent of an authority carrying out

torture) and the narrative of a torture situation we never experience (the ticking time-bomb scenario).

The malleability of authoritative intent allows for the switch from (a) the original intent of the torturers (assumed by Gary Bauer in the above quote) as leading to the *defusing* of the so-called 'ticking time bomb' – an intent that is replayed (as discussed below) in the popular *fictional* US program *24*, to (b) the quite different claim repeated by hosts of former Bush administration officials and 'conservative' US commentators following the killing of Osama bin Laden. Thus, how "back in 2003, 2004, and 2005, the mainstream argument for torture among its advocates was 'the ticking time bomb scenario.'" And yet, as Conor Friedersdorf continues, torture:

> apologists no longer feel the need to advocate for a narrow exception to prevent an American city from being nuked or a busload of children from dying. In the jubilation over getting bin Laden, they're instead employing this frightening standard: torture of multiple detainees is justified if it might produce a single useful nugget that, combined with lots of other intelligence, helps lead us to the secret location of the highest value terrorist leader many years later. It's suddenly the new baseline in our renewed national argument. *That's torture creep.*[27]

Indeed, it is torture creep, but put another way by Andrew Sullivan, torture idealists went from torture producing *true* and actionable information to torture being necessary to "break down terrorist suspects" so that they would *eventually* provide some bit of information vital – anywhere from six to eight years later – to an operation like the one which killed bin Laden. Sullivan adeptly characterizes this as "what we call creative reasoning."[28] Nevertheless, the connection between narrative extrapolation and faith (with regard to intent), means that what is at stake in support for torture is much more complex than an 'opinion' – it is a value, or a value system. When challenged, we are contending with not only a torture-supporter's opinion, but, more personally, their *faith* and the *legitimacy* of the person who shared with them a story about torture's probable efficacy. Thus, this might explain why 'people of faith' are most supportive of torture, and therefore why we might try to observe their discussions about torture in particular.

So, for students who find these arguments intriguing and would like to consider how to use them, I would argue that one consideration is that stories

don't just come from people we engage in our face-to-face everyday settings. They can and often are relayed through popular culture – including especially films and shows. The analysis of those 'sites' proves additionally useful in investigating the transition of torture narratives to our broader societies. Scholars, most notably William Connolly and Michael Shapiro, have developed micropolitics via **film analysis** for over a decade, and certain 'macropolitical' lessons about torture have been relayed through particular films as well.[29] You may find film analysis particularly useful for your own research – and it can be a fun way to pursue research as well (depending on the film, of course!).

I've utilized film analysis in my own work and on this particular topic. Consider the depictions of torture found in fictional representations like *24* but also the film *Zero Dark Thirty*, in that they are not only reflective, but *productive* of the narratives that can serve to justify its use. I explored this in a recent study.[30] What narratives emerge from the film? *Zero Dark Thirty* tells about the search for Osama bin Laden by the CIA as told through the perspective of one agent – Maya Lambert, played by Jessica Chastain. The movie suggests that torture, and specifically waterboarding, played a key role in revealing to the CIA the identity of bin Laden's courier, which then eventually led to bin Laden's location. Film analysis, as a micropolitical method, broadens out how narratives in a film relate to the ones proliferating in a broader political community. Indeed, the depiction in *Zero Dark Thirty* dovetailed with the assertions that had been issued by torture advocates from the Bush administration and CIA officials throughout the 2000s and especially following bin Laden's assassination. What is interesting about this particular narration is not its accuracy, but that it depicts a competent set of CIA officials gathering intelligence (accurate, we are made to believe) via torture. Further, the filmmakers dubbed the film a 'docudrama' and had been consulted by the CIA after the bin Laden raid. In a striking example of what James der Derian titled in 2001 the new 'military-industrial-media-entertainment network', the CIA agreed to cooperate with Mark Boal and Kathryn Bigelow (the movie's creators) because, according to CIA spokeswoman Marie Harf in one declassified email from June of 2011: "I know we don't pick favorites but it makes sense to get behind a winning horse … Mark [Boal] and Kathryn's movie is going to be the first and the biggest."[31] This example, largely fictional though it was, illustrates how important films and other practices of

'popular culture' are for 'teaching' political logics, including the 'logic' of torture.

A related third mechanism which might be producing torture's increasing popularity in the United States is something that I titled in one study the 'imaginative' strata of power: "Power's productive influence also depends upon connected individuals (in a nation-state these would be citizens), imagining that its operation is ethical and even beautiful."[32] Put another way, and again returning to the hiddenness of torture, we can't *see* torture when it is practiced, so we have to substitute some scenarios to imagine what it entails, what type of information it produces, and who both the tortured and torturers are and what their ethical 'makeup' is. This comes coupled with a more Manichean understanding of humans – some are good, some are evil, and never in the same person shall the twain meet. Such an emphasis on a faith in things we cannot see might prove instructive for understanding why the most religious Americans (and Gary Bauer, quoted above, qualifies here), express higher levels of support for torture than those who express their faith as 'unaffiliated' or as attending religious services 'seldom or never' in recent polls.[33] What also could be at work here is that those who see their 'faith' as particularly strong – that are willing to 'believe' that something exists that they can imagine but never see – are those who are more willing to *imagine* that the torture of individuals is both necessary and unproblematic. Second, such faith also lends itself to a belief in the righteous intent of *authority*, a trust that those who operate in the 'shadows' are doing so to physically secure our political communities from attacks. Thus, individuals of faith can *participate* in the torture of evil, imagining what the torture chamber looks like, who is doing the torturing (and who or what is being tortured) and what torture produces. Being able to imagine, on our own or with others, allows us to participate in the activity of torture – of defending the nation, or of personalizing the practice of torture ('what if your child was being held captive, what would YOU do'). It's all imagination – fantasy (for again as Rejali points out in his 2009 book, we know of no real 'ticking time bomb' scenario in all of human history) – but being able to imagine the torture scene makes it even more real for us.

What we might focus on when utilizing this mechanism and approaching our subjects would be how the *real images of torture*, such as when the Abu Ghraib photos from 2003 were made public in early 2004, might impact and challenge

this imagined scenario.[34] Again, the hiddenness of torture allows and even necessitates such imagination, a point I made in a previous study:

> if we do not see the subjects of disciplinary practices, if they are concealed, then we can imagine (by "forming a mental image" of them) that they are as bad as possible. The invisibility of the detainee is necessary for the citizen of the country detaining such individuals to perceive such treatment as necessary.[35]

Images of torture as it was *actually practiced* juxtapose with this imagination. While we cannot subject our subjects to such images, if they are familiar with those images, or even the narrated accounts by tortured victims (see above), we can see how respondents continue to rationalize torture when this imaginative scenario gets challenged and juxtaposed with accounts of torture itself.

Lessons to Be Learned

Micropolitical methods do have some limitations that researchers pursuing them need to be aware of – and 'good' micropolitical analysis usually has to resolve four sets of issues. A first limitation relates to access to our subjects. Even when we identify particular groups where torture is seemingly becoming more popular (such as the more religious, noted below), and want to hear what they have to say, it may be difficult to approach them. Issues over trust of the researcher may constrain subjects from talking about such topics, or generate skepticism and even anxiety over how their words will be used or represented in scholarly work.

For instance, about a decade ago I wanted to find out what US veterans thought about the topic of Just War, and wanted to carry out both surveys and participant observations at Legion Halls near my university at the time (the University of Kansas). In conducting a 'pilot study' I met with a focus group of veterans during an American Legion Hall dinner one evening. All were incredibly cordial and enjoyable to speak with, however, many expressed skepticism at what I would do with the 'results' of my conversations with them. One in particular, after finding out that I was not active or former military, expressed his concern that I would 'publish' my 'results' in the *New York Times,* or on *MSNBC.* I explained that academic

studies appear in peer-reviewed academic journals, and my goal was only to publish in those. Even after reassuring those in my focus group that their responses would remain anonymous, I could tell that their skepticism would remain. Discouraged, and due to funding and tenure-track issues that already constrained my timeline for research, I gave up the study altogether. The point that good micropolitical research is produced by *building trust* with respondents is important, and on a controversial topic like torture it can especially be a delicate process for gaining access to respondents.

A second, related issue is that the format of discussion may be too rigid or formal. This is often where micropolitical methods may depart, to some degree, from more rigid and formalized methods that social scientists pursue. In an ethnographic study, for instance, we may have particular questions we want to ask subjects, but if they are structured too rigidly it may inhibit the types of responses or conversations that could disclose some (if not all) of the narratives we find important. Cecelia Lynch discusses how this issue arose in her own research and how she resolved it in the following vignette:

> In my interviews of representatives of religious humanitarian organizations in Central and East Africa and the Middle East … I find that attempting an apolitical stance sometimes causes interviewees to treat the research questions less seriously and more perfunctorily, and the interview as a whole goes nowhere. Conversely, when I give reasons for the questions I pose, including relating aspects of my own background and experience, or acknowledging the political or cultural leanings which lead me to ask my questions, I find that interviewees are ready to engage and react, as long as I also begin from a basis of respect and let them know in so many words that I am there to learn from them to what degree my own assumptions can be supported or need revision.[36]

In short, the issues of both trust and generating 'data' with, and from, our subjects comes up quite frequently for scholars doing interpretive ethnographic work. Our research questions are formulated with one purpose, but that purpose might not 'match up' with what a subject wants to talk about. For critical and interpretive scholars, the importance of *agency* – our capacity to act in the world – must be considered not only in 'what' we are looking for in the world of politics but also as a value for our subjects as well (i.e., what they say, whether it leads to our own research outcomes or

not, matters). This issue relates to the first one, of course, in that the more conversational we are, the more likely we are to generate the *trust* with subjects that is so important for generating, in turn, rich narrative data that we want to analyze when it comes to a topic like torture's popularity.

Third, and especially considering the topic being discussed or evaluated, we also have institutional hurdles regarding human subject approval. One vignette might provide illustration here. At my previous institution, the University of Kansas, I helped serve as an advisor to a 'Global Scholar' project pursued by a talented undergraduate student. She was investigating how views on torture might be influenced by whether subjects were exposed to either the rhetorical representation of torture, or its image (like the Abu Ghraib photos). Even though what she was proposing would entail subjecting 'subjects' (her fellow students) to images they had likely already seen on the evening news, her proposal was rejected by the university's **Institutional Review Board**, the campus entity responsible for making sure that research is carried out ethically, includes informed consent and does not harm participants. Yet even this experience provided a very important story (and evidence, for that matter – for what does it say about torture if the images of it are too traumatizing to see?) when presenting her research at an end-of-the-year symposium.

Fourth, there is not only an analytically delicate, but also ethically sensitive, role played by the researcher of the topic that relates to the other three issues. This is a feature of the 'reflexivity' of critical and interpretative methods, a "practice that includes the ways in which the sciences themselves impact that which they study."[37] Such a reflexive awareness by the researcher doesn't entail qualifying everything about the relationship to the subject (and subjects) of a study, but it does require "making the relationship between the researcher and the researched explicit [to] create a space where the researcher can be conscious of their own understanding and meaning making, and allow for the reader to also be aware of these relationships"[38] or, again as Lynch notes above, showing 'respect' for our subjects and subject more broadly. So, in other words, who we are as researchers, regardless of how 'objective' we may wish to be, impacts how we relate to subjects (whether those are films, or narratives, or the people

we interview in our ethnographic methods), including the framing of the research questions used to explore the topic at hand. And, perhaps just as importantly, for some micropolitical methods (such as focus groups and/or ethnographic ones), it impacts how they relate to us as well. Yet, further, if our purpose is not only analytical but also *political* (i.e., if we want to not only access but perhaps challenge the 'stories' that sustain torture's popularity), then thorny issues over the role of the researcher in the moment of encountering our subjects emerges. Are we there to just record, or converse with our subjects? Or might we also try to have a dialogue with them so that they confront other possible ways to approach the topic?

If a scholar utilizing micropolitical methods handles these four issues effectively, they will likely have accessed a very rich set of narratives to better understand how people 'make sense' of politics generally. A successful study that utilizes this data is one that readers find persuasive, that is suggestive if not determinative of the types of arguments that have influenced (in the current case) **survey** respondents who support the use of torture. Such research calls our attention to both the analytical *and* political imperatives in both the problematic policies and practices of global politics (like torture), and also the ways in which research strategies and the questions we ask have a political nature to them. The hiddenness of torture itself means that strategies used to assess its increasing popularity disclose a social scientific purpose that is also critically normative. That is, most debates about torture via 'scenarios' or cinematic representations have themselves been quite anti-empirical, and so the more we can bring torture out into the 'open,' to discuss what it really is like or has been, what it historically has been useful for (such as confession or the demonstration of sovereign power) versus what it hasn't.

The empirics on torture are known but rarely shape public discourse. Darius Rejali's *Torture and Democracy,* over its 800-plus pages, debunks most torture myths that circulate in contemporary discourses.[39] And, on December 11, 2014, the US Senate Intelligence Committee released its report on the Central Intelligence Agency's (CIA) detention program of the 2000s. The release was accompanied by presentations on the Senate floor by the co-chairs of the committee, US Senators Dianne Feinstein (a Democrat from

California) and John McCain (2008 Republican presidential candidate from Arizona). Utilizing CIA documents, communications and papers, the report concluded succinctly and unequivocally in three (#'s 1, 2, and 10) of its 20 core findings that the 'enhanced interrogation' techniques were ineffective in acquiring intelligence from detainees.[40] The report methodically took apart the 20 most frequently cited claims by the CIA of plots disrupted, or captures of terrorist aides or suspects, directly linked (by the CIA) to their 'enhanced interrogation' program.

If the debates about torture were based in social scientific *fact* instead of philosophical counterfactual, there would be no debate. So, bringing more facts – however limited they may be – to the debates surrounding torture provides a nice opportunity for analytical juxtaposition as well as critical purposes. Yet, and here's the rub: one of those (qualified) facts also includes the increasing popularity of torture in the United States – in spite of torture's known inefficiencies, both moral and strategic.

As a result, the benefits of all critical methodologies are that they *enrich* the suggestive data at our disposal, but they also bring us *closer to our subjects* in ways that can prove heuristically important as well. That is, what other issues seem important to citizens in a democracy? What animates their understandings of world politics, and where do they get those understandings in the first place? This chapter has provided some directions for seeking out what remains largely hidden from scholars and calls on us to perform such seeking, even with all its limitations, in a world that seems to be increasingly disorderly, surprising and difficult to grasp from afar.

Interested to Know More about the Study Discussed in this Chapter?

Steele, Brent. 2013. "The Insecurity of America: The Curious Case of Torture's Escalating Popularity." In *Justice, Sustainability, and Security: Global Ethics for the, 21st Century*, ed. Eric Heinze. New York: Palgrave,

171–204.

Exercises and Discussion Questions:

1. What do you see as the primary strengths of micropolitical approaches to understanding torture's popularity, and such approaches generally? What are the primary weaknesses of such an approach?
2. How might we combine one of the other methods discussed and developed in one of the other chapters of this textbook with the micropolitical approach the author proposes here?
3. Since this approach highlights some of the ethical issues with micropolitical research, what do you think are the best ways to train scholars to pursue them to foster, not only methodologically and analytically, but also ethically, sound research vis-à-vis subjects? Might the discussion in Chapter 11 on Institutional Reviews Boards and experimental design be of use? Why or why not?

Recommended Resources:

Center for Victims of Torture (https://www.cvt.org/resources/researchers): a St. Paul-based global NGO that not only collects narratives of victims of torture, but also provides a checklist for doing ethnographic work with them.

Notes

1 Steele 2008; 2010; 2017; 2013.

2 Foucault 1977.

3 Connolly 2002.

4 Rejali 2009; Gronke *et al.* 2010

5 Gronke *et al.* 2010: 437.

6 Malka and Soto 2011, whose study I briefly return to below, take a 2004 survey and a 2008 survey – but other than citing them in passing, ignore the surveys taken over the past two years which show a clear uptick in torture's popularity among the American public.

7 Gronke *et al.* 2010: 439.

8 "Survey on International Humanitarian Law." March 2011. www.redcross.org/www-files/Documents/pdf/international/IHL/IHLSurvey.pdf

9 See "US seen as less important," http://people-press.org/2009/12/03/us-seen-as-less-important-china-as-more-powerful/, Pew, December 3, 2009, and "The Religious dimensions of the torture debate," Pew, April 29, 2009 http://pewforum.org/docs/?DocID=156

10 In fact, the reservations the United States made in ratifying the Convention in 1994 were to use definitions of "cruel, unusual and inhumane treatment or punishment" set forth in the US Constitution. See U.S. Code, 2340, 2340a and 2340b, www.law.cornell.edu/uscode/18/usc_sec_18_00002340---A000-.html

11 This is the figure for those who attend 'at least weekly'; for those who attend 'monthly or a few times a year' it is 51 percent (74%).

12 The overall sample was 724 respondents. The breakdown categories of religion range between 94 and 174 each). http://pewforum.org/docs/?DocID=156

13 Malka and Soto 2011: 109.

14 Gronke *et al.* 2010: 438.

15 Allahpundit, "Pew Poll: public support for torture at five-year high," http://hotair.com/archives/2009/12/04/pew-poll-public-support-for-torture-at-five-year-high/

16 Goldman and Craighill 2014.

17 53 percent even agreed that torture "produced important information that could not have been obtained any other way" (Goldman and Craighill 2014).

18 Wendt 1998.

19 Solomon and Steele 2017: 270.

20 Walsh 2012.

21 Walsh 2012: 517.

22 Stevens and Vaughan-Williams 2016.

23 Stevens and Vaughan-Williams 2016: 10.

24 https://www.cvt.org/resources/researchers

25 Arendt 2013: 179.

26 "Just War Theory and Enhanced Interrogation." http://townhall.com/columnists/GaryBauer/2009/08/28/just_war_theory_and_enhanced_interrogation_ho

27 "Torture Opponents were Right," *The Atlantic*, May 5, 2011. www.theatlantic.com/politics/archive/2011/05/torture-opponents-were-right/238387/

28 "Quote for the Day," Andrew Sullivan, May 5, 2011. http://andrewsullivan.thedailybeast.com/2011/05/quote-1.html

29 Connolly 2002; Shapiro 2011.

30 Steele 2017.

31 Tau 2012.

32 Steele 2010.

33 "Polls show support for torture among Southern evangelicals." http://pewforum.org/news/rss.php?NewsID=16465, and "The Religious Dimensions of the Torture Debate." http://pewforum.org/docs/?DocID=156

34 Steele 2010.

35 Steele 2010: 153.

36 Lynch 2008: 718.

37 Amoureux and Steele 2016: 10.

38 Dottolo and Tillery 2016: 129–130.

39 Rejali 2009.

40 US Senate 2015: 3–12.

References

Amoureux, Jack, and Brent J. Steele, eds. 2016. *Reflexivity in International Relations.* London: Routledge Arendt, Hannah. 2013. *The Human Condition.* Chicago: University of Chicago Press.

Connolly, William E. 2002. *Neuropolitics: Thinking, Culture, Speed.* Minneapolis, MN: University of Minnesota Press.

Dottolo, Andrea, and Sarah Tillery. 2016. "Reflexivity and Research: Feminist Interventions and their Practical Implications." In *Reflexivity in International Relations,* ed. J. Amoureux and B. Steele. London: Routledge, 123–141.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison.* New York: Pantheon Books.

Goldman, Adam, and Peyton Craighill. 2014. "New Poll Finds Majority of Americans Think Torture was Justified after 9/11 Attacks." *Washington Post,* December 16, 2014.

Gronke, Paul, Darius Rejali, Dustin Drenguis, James Hicks, Peter Miller, and Bryan Nakayama. 2010. "US Public Opinion on Torture, 2001–2009." *PS: Political Science & Politics* 43(3): 437–444.

Lynch, Cecelia. 2008. "Reflexivity in Research on Civil Society: Constructivist Perspectives." *International Studies Review* 10: 708–721.

Malka, Ariel, and Christopher J. Soto. 2011. "The Conflicting Influences of Religiosity on Attitude Toward Torture." *Personality and Social Psychology Bulletin* 17(8): 1091–1103, 1092.

Rejali, Darius. 2009. *Torture and Democracy.* Princeton, NJ: Princeton University Press.

Shapiro, Michael J. 2011. "The Micropolitics of Justice: Language, Sense, Space." *Law, Culture and the Humanities* 8(3): 466–484.

Solomon, Ty, and Brent J. Steele. 2017. "Micro-Moves in International Relations Theory." *European Journal of International Relations* 23(2): 267–291.

Steele, Brent J. 2008. "Ideals That Were Never Really in Our Possession: Torture, Honor and US Identity." *International Relations* 22(2): 243–361.

Steele, Brent J. 2010. *Defacing Power: The Aesthetics of Insecurity in Global Politics.* Ann Arbor, MI: University of Michigan Press.

Steele, Brent J. 2013. "The Insecurity of America: The Curious Case of Torture's Escalating Popularity." In *Justice, Sustainability, and Security: Global Ethics for the 21st Century,* ed. Eric A. Heinze. London: Palgrave.

Steele, Brent J. 2017. "Organizational Processes and Ontological (in)Security: Torture, the CIA and the United States." *Cooperation and Conflict* 52(1): 69–89.

Stevens, Daniel, and Nick Vaughan-Williams. 2016. *Everyday Security Threats: Perceptions, Experiences, and Consequences.* Oxford, UK: Oxford University Press.

Tau, Byron. 2012 (August 28) "Emails: Obama Admin. Wanted to Back 'Winning' Movie." *Politico44 Blog*, https://www.politico.com/blogs/politico44/2012/08/emails-obama-admin-wanted-to-back-winning-bin-laden-movie-133492 (last accessed June 22, 2018).

United States Senate Intelligence Committee. 2015. *The Senate Intelligence Committee Report on Torture.* London: Melville House.

Walsh, Katherine C. 2012. "Putting Inequality in Its Place: Rural Consciousness and the Power of Perspective." *American Political Science Review* 106(3): 517–532.

Wendt, Alexander. 1998. "On Constitution and Causation in International Relations." *Review of International Relations* 24(5): 101–118.

Statistical Research

Lack of Citizenship, the Achilles' Heel of Latino Political Power

Adrian D. Pantoja and Sarah Allen Gershon

CONTENTS

Getting Curious: Did Latino Voters Live up to the Hype in 2016?

Much was made of the critical role Latino voters would play during the 2016 election. Essentially labeled "Hillary Clinton's firewall" by the media and pundits,[1] the growth of the Latino electorate, coupled with a strong preference for Hillary Clinton among Latinos, all but guaranteed that she would win the presidency. While nationally Latinos are about 16 percent of the electorate, their large presence in key battleground states make them pivotal voters. As students of history, politics and the Latino electorate, we believed that Donald Trump had made a strategic miscalculation by attacking immigrants, Latinos and other minority groups. He did not hold back in expressing his animosity toward Mexico and unauthorized persons from that country when he announced his candidacy on June 16, 2016 with these words, "When Mexico sends its people, they're not sending their best. They're not sending you. They're not sending you. They're sending people that have lots of problems, and they're bringing those problems with us. They're bringing drugs. They're bringing crime. They're rapists. And some, I assume, are good people." Donald Trump made illegal immigration and Latinos scapegoats for what he perceived were the nation's ills. His rhetoric and policy positions on immigration were reminiscent of that which was experienced in California in the 1990s.[2] In California, attacks against immigrants and Latinos in the mid-1990s led Hispanics to turn the Golden State solidly blue.[3] In 2016, Latino voters were expected to turn out in record numbers against Donald Trump.

For eight weeks prior to the 2016 presidential election, Adrian Pantoja (the first author of this chapter) was running a weekly tracking poll of the Latino electorate for the polling firm Latino Decisions. The tracking poll allowed scholars to project the voting behavior of Latinos on Election Day. Let's examine what the tracking poll revealed to illustrate why Democrats were extremely optimistic on Election Day. We found that Latinos were poised to turn out at very high rates. In week three of the poll, 71 percent of Latino registered voters said they were "almost certain" to vote on Election Day. By week 8, that number had climbed to 80 percent. Over 50 percent of Latino voters said they were more enthusiastic about the 2016 Election than they were about the 2012 Election. Over three-quarters of Latinos rated Donald Trump "unfavorably." Close to half of respondents said the Republican Party was "being hostile" towards Latinos. Over three-quarters said they were certain about voting for Clinton. These and other

polling results led us to project that Hillary Clinton would win about 79 percent of the Latino vote, while Donald Trump would win a mere 18 percent of their vote.[4] Based on this data, we expected that Latinos would help pave the road to Clinton's victory in 2016. In fact, Professor Pantoja wrote a blog post on Election Day titled, "Latino Voters Will Block Trump's Path to the White House" (Pantoja 2016). We were wrong.

Immediately after the election, the media and pundits were scrambling to understand how Donald Trump prevailed over Hillary Clinton. Some of that focus turned to Latino voters. It turns out that Clinton did win nearly 79 percent of the Latino vote as projected by Latino Decisions. Yet, their numbers were not sufficient to overcome those who voted for Trump. This led some to ask if the power of the Latino vote in 2016 was overhyped. Was Hillary Clinton's firewall merely myth? Some experts on Latino politics, like Professor Rodolfo O. de la Garza from Columbia University, would contend that the power of the Latino vote has been overstated in *every* presidential election since 1960 when John F. Kennedy made an effort to mobilize Mexican American voters in Texas. Then and now, activists stressed the power of the Latino vote in delivering the presidency. The facts are that the Latino population continues to grow exponentially every decade. In 1960 there were 6.3 million persons in the United States of Hispanic ancestry. Today, there are close to 60 million Latinos in the country.[5] The Latino presence is undeniable. While it is true that demographic growth has increased Latino voting strength, the fact remains that Latino electoral power significantly lags behind their demographic heft, earning them the name "the sleeping Giant."[6] In 2016, of the 60 million Latinos, a mere 12 million voted. Of course, not all 60 million are eligible to vote. Two of the biggest barriers limiting Latino political power are: 1) having large numbers who are under the age of 18 and 2) having a sizable population who are non-citizens. The former will be remedied by time, while the latter has been the subject of intense debate and analysis. Until Latinos address their low rates of naturalization, their political power in the US will remain limited. In 2006, we conducted research designed to understand this problem of why some Latinos and Latinas move towards naturalization (allowing them to vote in US elections) and why others remain non-citizens. In this chapter, we outline our research design, analysis and conclusions. While ten years have passed since we conducted our research, the

2016 Election illustrates that the research question we examined remains relevant today.

Writing a research paper on why Latinos do or do not pursue naturalization came out of a series of discussions on this subject in a graduate seminar on Latino Politics between Professor Pantoja and his then graduate student Sarah Allen Gershon (now a professor at Georgia State University). Professor Pantoja had assigned a series of articles on naturalization; most of which emphasized the importance of length of residency and socio-demographic characteristics. Both of us hypothesized that other factors were important, such as having an interest in politics and attitudes about the value of voting. These political orientations are found to shape political participation among citizens. Therefore, we surmised that similar orientations would shape the naturalization decision among non-citizens. However, this was just a **hypothesis** (an educated guess), and it remained to be seen whether there was data available that could help us test that hypothesis.

Since the question of why Latinos pursue or forgo the naturalization process is both theoretically and politically important, we decided to collaborate in an effort to answer this question. The **research question** is the question researchers seek to answer in their academic writings through the gathering of evidence or data. Almost every political science journal article, book chapter or book is written to answer a question or series of questions. Typically, a single question is sufficient in order to make a research project manageable. In our paper, we were broadly concerned with the following question: What are the factors that promote or impede the pursuit of naturalization among Latino immigrants?

As noted earlier, political scientists like to answer questions that are both theoretically and politically important. We believed our question met these criteria. The political importance of the question is obvious, given that the size of the Latino electorate could double if its non-citizen population became US citizens. Having a sizable electorate will not only increase Latinos' political clout but will also shape the American political landscape. Hence, the findings from our project have the potential to inform policymakers and activists seeking to increase Latino political power. Theoretical importance simply means academically important. In other words, do academics in your field, find your question interesting? More importantly, will your answer to this question add to the existing academic literature on this subject? If the answer is yes to both, then

you are well on your way to writing a theoretically important paper.

Coming up with a good research question is both fun and challenging, and it has been the case for both of us that some of the best ideas (research questions) can come over a cup of coffee at Starbucks or a round of beers at a pub. The times and places for ideas are never ending. Of course, the challenge is making sure you are asking a question that has not already been asked; finding an answer that has not been developed by others; or using a method, measurement or **data** that has not been tried before. In short, the whole purpose of research is to come up with something new. There is nothing innovative about research that simply re-states what someone else has said or found. We believe it is far more creative, rewarding and fun to develop original research. And yes, we believe research is fun, otherwise, we would not be doing it!

The Research Strategy: What is Statistical Analysis?

Political scientists, like other social scientists, have a number of methodological tools that allow them to test whether a particular hypothesis is true or false. Throughout this book, each author is highlighting the advantages and disadvantages of the method they use. Both of us are trained in a variety of methods, yet our method of choice (due largely to our interest in mass political behavior) is statistical analysis. **Statistical analysis** is often used when working with large numbers of observations, allowing researchers to analyze data in a systematic and convenient way. For example, if we examine the likelihood of a single individual immigrant to naturalize, then statistical analysis would not be necessary (or effective); however, with a large-scale survey (such as the one used here), statistical analysis can help researchers make sense of hundreds of observations. **Surveys**, asking many **closed-ended questions** to large numbers of people (see [Chapter 8](#)), allow researchers to take those responses and quantify them (essentially assigning numerical values to responses).

Statistics allow us to make sense of responses on large surveys like the one used in our paper. There are two broad categories of statistics: **descriptive statistics** and **inferential statistics**.[7] **Descriptive statistics** allow us to identify

various characteristics of large groups. For example, using descriptive statistics we are able to identify differences in naturalization among male and female immigrants in our sample. **Inferential statistics** allow us to make inferences about a number of things based on our results, including the relationships between different variables, and the generalizability of sample results to a **population**.[8] Inferential statistics allow us to determine whether that relationship is **statistically significant**, meaning that it likely exists in the population. For example, using inferential statistics, we can employ sample data to draw conclusions about the impact of various political orientations on immigrant naturalization in the population. For our paper, we relied primarily on inferential statistics because we wanted to use our sample data to draw conclusions about Latino immigrants in the US.

Issues with this Research Strategy

When are statistical analyses appropriate? What are the advantages and disadvantages of this methodology? The primary advantage of working with statistics is that you can generalize the findings (based on a representative sample) to a larger population (a process referred to as **statistical inference**). For example, our study analyzes the thoughts and behaviors of 1,042 Latino immigrants (our sample); but through the use of statistical analysis, we are able to make inferences about the thoughts and behaviors of 12.8 million immigrant Latinos (the population of immigrant Latinos reported by the 2000 US Census).

One of the issues that researchers relying on non-experimental data often confront is the difficulty associated with isolating causal relationships. When trying to examine a causal relationship using non-experimental methods, a common problem researchers run into is their ability to account for multiple causes of the dependent variable. Essentially, for us to be able to assert that one variable *causes* a change in another, we must be able to control for the impact of other variables (which may actually be responsible for the observed change). One of the most commonly used examples for the need to eliminate alternative explanations is the observation that both crime levels and ice cream sales tend to

increase in the summer. While these variables might be correlated with one another, that does not necessarily mean one is *causing* a change in the other. Instead, it is likely that some other third variable (e.g., warm weather) is responsible for a change in both crime and ice cream sales. Thus, for us to examine whether political orientations are *causing* a change in naturalization rates among Latino immigrants, we have to deal with other variables that likely impact naturalization rates as well.

Clearly, we live in a multivariate world, and it is always a challenge to identify *all* of the factors that explain political phenomena, or any phenomena for that matter. As a consequence, some scholars have greater confidence in findings drawn from experimental designs (see [Chapter 11](#) for a discussion of experimental designs). Due to their high levels of control, experiments allow researchers to identify with greater precision and confidence the impact of one variable on another. While experiments provide the best test of causality, they are often limited in their **external validity** (or ability to generalize findings to larger populations), and are not well suited for every research question, particularly those interested in making inferences about a large population. When researchers are interested in understanding the relationship between variables for a large population, they can rely on statistical analysis to help deal with this problem, statistically 'controlling' for the impact of other variables by relying on multivariate analyses (as we do in this chapter). These analyses allow us to examine the impact of one variable on another, holding the impact of other variables constant. By statistically accounting (or controlling) for other variables which may cause variation in the dependent variable, researchers may be more confident about the relationship they observe between their key independent and dependent variable.

While statistical analysis may help us make inferences about large populations based on relatively limited samples, and rule out alternative causes of the dependent variable, it does have some limitations. First, the use of statistical analysis (based on mathematical measures) may not allow for the kind of nuanced understanding of the subject that qualitative research might generate. Relying on aggregate summaries of large quantities of data may limit our ability to recognize individual differences in the respondents' understanding of the questions and their responses. In fact, we had to rely on qualitative research

conducted by other scholars[9] to help us interpret the results of our statistical analysis because of the limitations of this type of data and method. Furthermore, examining a large number of responses together may lead us to overlook critical outliers – respondents that are very different from the mean values. We expect random variation in the data to exist, but unusual cases may tell an important part of the story that will be missed by a large-scale statistical analysis.

The Literature Review

How do we know our research question is interesting or that our approach to answering it is original? We gather and read research published by scholars on the topic we are planning to investigate. This process of gathering, reading and summarizing research on the topic we are investigating is known as the **literature review**. Typically, scholars research topics where they have accumulated some level of expertise. This expertise comes from reading works on the topic and adding to this literature by doing original research. Since one of the areas we specialize in is Latino politics, both of us have come across literature on naturalization in graduate courses or through our own research. It is unusual for scholars to carry out research on a subject that is completely new to them.

We both employed a variety of library databases and typed in key words such as "naturalization," "citizenship," "Latinos," "immigrants," "political incorporation," "civic engagement," and a few others. Since we were developing a scholarly article to be published in a **peer-reviewed** or **refereed journal**, we limited our search to articles published in such journals and academic books as opposed to articles in popular journals, magazines or newspapers. That's because the latter do not go through a blind **peer-review** process. Peer-review, blind-review, or refereed journals are terms describing the process an article undergoes before it is published in an academic journal. Peer-reviewed journals send submitted articles to two or more scholars in the field for review. The authors of the article do not know who is reviewing their work and the reviewers do not know who wrote the article (since the names are removed from the cover page). The process ensures that those reading the article base their evaluations, or

review, on the merits of the work and not personal ties to the author(s). Based on the reviews, the editor of the journal will recommend that the article be published, revised and resubmitted, or rejected. The process is long and rigorous, but it ensures that only top-quality works meeting the journal's standards are published. Works published on the internet, or non-academic journals, books, magazines and newspapers typically do not have the same type of peer-review process.

After getting many hits from the academic databases and printing hard copies of the articles, we looked at their reference pages to see if they referenced articles that were not on our list. The next task was to read these articles and organize them into several themes. The purpose of literature reviews is not to summarize the findings of every single article or book written on a topic. Given that we had over 20 articles on the subject of naturalization, summarizing each article in a single paragraph would have created a journal article with a literature review over ten pages in length! In addition, simply summarizing each article does not explain how they fit together. Literature reviews usually constitute about 10 percent of an article.

As noted earlier, we hypothesized that having certain political orientations, namely a positive outlook toward voting and interest in politics, would spur naturalization among Hispanic immigrants. From the literature, we developed a definition of political orientations to be "citizens' subjective feelings about the political system: whether they know and care about politics, desire to participate in politics, and feel capable of affecting change in the political system."[10] The literature review also allowed us to identify if the connection between political orientations and naturalization had already been made by other scholars. We found that the literature on naturalization could be divided into four themes. Each theme represented research that highlighted the importance of certain **independent variables**, the variables that predict or *cause* the dependent variable to occur. For example, we believe that naturalization (our **dependent variable**) is *caused* by (1) socio-demographic, (2) cultural, (3) contextual, and (4) transnational variables. Socio-demographic variables include an individual's level of income, education, marital status and age. Cultural factors include measures of English proficiency, length of residency and ancestry group (e.g., Mexican, Cuban, and so forth). Contextual factors capture immigrants' reasons for migration, experiences

with discrimination and the characteristics of the source country, such as its regime type and level of development. Finally, transnational factors are the ties immigrants maintain with the source country while residing in the United States.

The literature review identified a number of **hypotheses** about why immigrants pursue or do not pursue naturalization. It identified the nature of the relationships (e.g., positive, negative or non-existent) between the independent variables and the dependent variable. It also identified how these variables were **conceptualized** (defined) and **operationalized** (measured), as well as the type of data sources used. Armed with all of this information, our task, then, was to carry out the research and make a contribution to the literature by doing something different or original.

We were surprised to find that none of the previous works on immigrant naturalization considered the role of political orientations in shaping the decision to pursue citizenship. This was a gap in the literature we wanted to fill. Perhaps this gap was based on the belief that immigrants did not become US citizens because they desired to participate in politics. Rather, it was believed that most became US citizens out of a desire to increase their economic mobility or to sponsor family members for legal residency. We set out to show that the conventional wisdom was incomplete, and that many immigrants were pursuing naturalization out of a desire to participate more fully in the American political system.

The semester we began our project, one of us (Adrian Pantoja) had also assigned articles on gender and politics in his Latino Politics graduate seminar. One of the key themes from this literature is that Latinas (women) conceptualize and participate in politics differently than Latinos (men). Therefore, we believed that the naturalization decision is likely to vary by gender. In other words, we were not merely interested in determining whether Latino immigrant men and women had different rates of naturalization, but whether the factors underlying that decision differed by gender.

We noted in our paper that "the existing quantitative scholarship on immigrant naturalization is lax when it comes to understanding the extent and sources of gender differences in citizenship acquisition. Many studies simply fail to control for gender in the multivariate models."[11] In addition, the effect of gender on naturalization is contradictory, with some finding women to be more likely to

naturalize and others finding the opposite, leading us to write, "The dearth of conclusive evidence over the impact of gender on citizenship acquisition highlights the need for gendered approaches to the study of immigrant naturalization."[12] We had two key hypotheses that had never been examined before in the literature:

> Hypothesis 1: Immigrants with positive political orientations are more likely to naturalize;
> Hypothesis 2: The factors structuring the naturalization decision are going to vary by gender.

In the first hypothesis, our independent variables are positive political orientations, while in the second hypothesis our primary independent variable is gender. Throughout the analyses, our dependent variable is naturalization.

Conducting the Study: Collecting the Data and doing the Analysis

To test our hypotheses, we relied on data from the 1999 "National Survey on Latinos in America," a survey carried out by telephone and sponsored by *The Washington Post*, Henry J. Kaiser Foundation and Harvard University. We used this survey because at the time it was one of the only surveys with a nationally representative sample of Latinos. The survey had a sample size (*n*) of 2,417 Latinos. Since our focus was immigrants, or foreign-born Latinos, we needed to remove some of the respondents from the full sample. We limited our analysis to foreign-born Latinos who were either naturalized citizens or eligible for naturalization. Thus, we eliminated respondents who were born in the US, Puerto Ricans (since they are US citizens by birth), or immigrants who have resided in the US for less than five years (the required waiting period before applying for citizenship). We also dropped individuals from our sample who were ineligible for citizenship because they were likely undocumented immigrants. We identified

undocumented respondents through a question that asked why they were not pursuing US citizenship. After removing these respondents, the final sample or *n* used in our analysis was 1,042 immigrant Latinos.

Depending on the nature of the concept, variables may fall into one of four levels of measurement: nominal, ordinal, interval or ratio. **Nominal** (also called categorical) **variables** are variables which have unordered categories (e.g., religion or gender). **Interval level variables** possess categories which are equidistant to one another, but the variable has no true zero (e.g., temperature). **Ratio level variables** possess full mathematical properties (e.g., age, income). Our dependent variable (Latino immigrant naturalization), is measured using an **ordinal measure**. Ordinal measures capture variables which have ordered (but not necessarily equally spaced) categories. For example, survey questions often present respondents with a set of responses to choose from which are ordered, such as (3) *strongly agree*, (2) *somewhat agree*, (1) *somewhat disagree*, and (0) *strongly disagree.* While it is clear that a respondent who "strongly agrees" with something has a higher level of agreement than one who "somewhat disagrees," the difference in these attitudes is not mathematically precise. That is to say, differences in ordinal measures vary by intensity or strength, but the exact magnitude of the differences in the categories is unknown. To illustrate this point, we could say that the difference between 1 pound and 3 pounds is 2 pounds, or the difference between 3 miles and 5 miles is 2 miles. These differences in ratio-level variables are mathematically meaningful and precise. Conversely, it is unclear what the distance (or difference) is between ordinal categories such as "strongly agree" and "somewhat agree."[13]

In this case, the ordered categories represent immigrant's *progress* towards citizenship. Specifically, we examine whether immigrants have: no plans to apply for citizenship (coded as 0), are planning to apply for citizenship (coded as 1), were in the process of becoming a citizen at the time of the survey (coded as 2), or were already naturalized citizens (coded as 3). The ordered measure (0, 1, 2, 3) is preferable to a **dichotomous measure** (0 "not a citizen"; 1 "a citizen"), since we believe that there is a meaningful difference between individuals who have no plans to become citizens, versus those who are planning to initiate the process, undergoing the process or have become US citizens. To ensure that there was variance on our dependent variable, our first step was to examine how the survey

respondents were distributed in their progress towards citizenship. Without variance in the dependent variable, there is little point in proceeding with our analysis since the study was concerned with why some immigrants naturalize while others do not. To examine the distribution of the dependent variable, we generated some descriptive statistics, which tell us something about the distribution or central tendency of a single variable – in this case, the proportion of respondents that fall into each category of the dependent variable. The results demonstrated that about half of the respondents were already citizens, approximately 16 percent were in the process of applying for citizenship, almost 26 percent were planning to apply, and a little less than 9 percent had no plans to apply.

Previous scholarship has demonstrated that immigrants' nation of origin sometimes affects the speed at which they become US citizens. We examined differences in the dependent variable by nation of origin or ancestry groups (e.g., Mexican, Cuban, Salvadoran and other Latinos), relying on **analysis of variance (ANOVA)** tests. ANOVA essentially allows us to see whether the groups we are examining are distinct through a comparison of mean naturalization rates among these different ancestry groups. Remember that when we rely on sample data to make some claim about a population characteristic or value, we are engaging in statistical inference. When examining samples, we often rely on statistical tests for significance to determine whether to reject the **null hypothesis** being tested. The null hypothesis, simply stated, is that there is *no* difference among various groups in the population, or that there is *no* relationship between independent and dependent variables in the population. In this example, we are using ANOVA tests to examine the probability that the null hypothesis (in this case, that average rates of Latino naturalization among groups from different countries are the same in the population) is true.[14]

The results demonstrate statistically significant differences in naturalization progress among the different groups. Specifically, Cubans had the highest rate of citizenship acquisition among the three groups, with 75 percent being US citizens. Salvadorans, on the other hand, had the lowest rate of naturalization, 22 percent were US citizens. Mexican respondents, at 36 percent being US citizens, fell between these two extremes. When we say that the difference in means among the groups in our sample is statistically significant, we are saying that the

probability (p) that the null hypothesis is correct is quite low. In the field of political science, we often see scholars reporting multiple p-values in their articles to let the readers know what level of confidence they have regarding their results. The most common p-values reported are .10, .05 and .01. We can interpret these numbers as the probability that the null hypothesis is correct; for example, p-values of .10, .05, and .01 indicate a 10, 5, or 1 percent chance (respectively) that the null hypothesis is correct. In this case, our results indicated that the probability (p) that naturalization rates among these groups are actually the same is less than .05 (or less than 5%). Thus, we are fairly confident that the differences we observed in our sample are representative of differences in the population at large.

While the results of the ANOVA tests suggest that nation of origin might explain some of the variance in the dependent variable, there are a number of other independent variables which may explain Latino immigrants' choice to naturalize. Until this point, we had yet to test competing hypotheses explaining naturalization. For the remainder of this analysis, we attempt to statistically identify the relationship between our independent variable of interest – political orientations – and naturalization (our dependent variable), while statistically accounting for alternative explanations of naturalization (e.g., nation of origin, socioeconomic status, transnationalism). Through the use of **multivariate analysis**, we are able to examine the impact of these competing variables on naturalization simultaneously. In the section "Testing the Hypotheses: The Results" below, we will explain what we mean by multivariate analysis.

Our primary independent variable – political orientations – is operationalized (or measured) using the responses to two different questions. The first is based on a question asking the degree to which an individual agrees or disagrees with the statement: "Voting is a waste of time." The variable, Voting is a Waste, is based on a four-point scale, ranging from 0 "disagree strongly" to 3 "agree strongly." The second variable is based on a question about political interest: "How much attention would you say you pay to politics and government?" The variable is based on a four-point scale ranging from 0 "none at all" to 3 "a lot."

While political orientations are our primary variables of interest, in order to assert that they are related to naturalization, we must control for the impact of alternate explanations. In other words, to be sure the relationship we identify is

real, we have to (statistically) account for other variables which may explain Latino immigrants' choice to naturalize. How do we know which variables to include in our model? This decision is based on theoretical reasoning and previous scholarship. Recall that our review of the previous literature identified a host of different variables which impact naturalization, including respondent's nation of origin, experiences with discrimination, reasons for migration, transnational ties, age, education, income, gender, marital status, length of residency and English-language proficiency. In our analysis, we attempt to statistically account or "control" for the impact of these predictors in order to clearly identify the independent impact of our variables of interest on naturalization.

Testing the Hypotheses: The Results

There are many different types of multivariate **models** which researchers use to examine the relationship between multiple variables. In this case, we chose to rely on **ordered logistic regression** (often call ologit for short). Ordered logistic regression is commonly used in large *n* analyses when the dependent variable is ordinal. Essentially, the ordered logistic regression estimates the impact the independent variables have on the *likelihood* of observations being in the next highest category of the dependent variable (in this case, progress towards naturalization).[15]

Table 7.1 Ordered Logistic Regressions Predicting Latino Immigrants' Progress towards Immigration (truncated)

	(Full Sample)		(Men Only)		(Women Only)	
	Coefficients (S.E.)	Min-Max	Coefficients (S.E.)	Min-Max	Coefficients (S.E.)	Min-Max
Voting Is a Waste	−.124(.061)*	−0.076	−.076(.087)	−.048	−.159(.087)*	−.095
Political Interest	.134(.075)*	.079	.220(.104)*	.135	.026(.109)	.015
Gender (Female)	.229(.136)*	.045	---	---	---	---
Sample Size	997		487		510	

K: *p<.05, **p<.01

Table 7.1 displays a portion of our findings – specifically, the impact of political orientations for the full sample (male and female respondents), for

Latinos (male respondents) only, and for Latinas (female respondents) only. Before discussing our results, let's review what the numbers in the table mean. Keep in mind that the statistical concepts described are typically learned in an advanced statistics course, so don't be concerned if the concepts may not be entirely clear at the moment. They key point is that the numbers on the table are mathematical properties describing (1) whether the variables are positive, negative or unrelated to the dependent variable; (2) whether the impact of the independent variables on the dependent variable is statistically significant; and (3) whether the impact or effect of the independent variables on the dependent variable is large, medium, small or negligible. Nonetheless, we will use common terms used by statisticians to describe these three points. The table displays two sets of results for each model. First, we see the **unstandardized coefficients** with the **standard error** in parentheses. The coefficients tell us the impact of each independent variable on the log odds of respondents' progress towards naturalization. Specifically, the coefficient gives us the expected increase in respondent's log odds of being in a higher naturalization category for every one unit change in the independent variable, controlling for all other variables in the model. For example, in [Table 7.1](#) (full sample), the negative coefficient for Voting is a Waste (-0.124) tells us that a one unit increase in respondents' belief that voting is a waste of time *decreases* their log odds of being in a higher naturalization category by .124, holding all other variables in the model constant. To assist with our understanding of the relationships between these variables, we present a statistic called "min-max" (discussed later on) which helps us understand the relative impact of the independent variables on the dependent variable.

The standard errors for the individual coefficients are also presented. When we discuss standard errors, essentially we are talking about the accuracy of our estimates, given that we are only using a sample of the population we are trying to generalize to. In order to draw inferences about the population (in this case, Latino immigrants), relying on a sample of approximately 1,000, we have to account for the differences between our sample and what the true value of the population may be. Remember, each sample selected will vary slightly from what we would expect to see if we drew a different sample, or if we examined the entire population. The standard error captures what we call the standard

deviation of the sampling distribution. In other words, the standard error captures the difference that will occur randomly between any sample and the entire population. Generally speaking, the larger the sample (and thus the greater proportion of the population being captured), the smaller the standard error is.

Notice that several of the coefficients have stars (e.g., *) after them. The key below the table tells us that these stars represent the **p-value** of the coefficient. The p-value refers to the statistical significance of the relationship in the multivariate model. In this instance, a p-value of .05 or below gives us a high level of confidence that the relationship we're observing is not coincidental. In these instances, we refer to the exhibited relationship as "statistically significant."

Aside from the coefficients, the standard errors and p-values, perhaps the most important value is the "min-max" which helps us interpret the effect of the independent variables on the dependent variables. Again we want to know if the effect is large, medium, small or negligible. This measure allows us to make standardized comparisons about the relative impact of the different independent variables on the dependent variable. More specifically, this statistic estimates the change in predicted probabilities of the dependent variable, given a fixed change in the independent variable from its minimum to maximum value, holding all other variables constant at their means.[16] So, for example, in the full model, the min-max value for gender (female) is .045. This number tells us that, holding all other variables in the model constant at their means, being female (rather than male) increases the respondents' probability of moving closer to naturalization by .045.

Now let's go over the results in Table 7.1. First, we examine the impact of political orientations on Latino immigrants' progress towards naturalization, using the entire sample (men and women). Both of our measures are statistically significant predictors of naturalization Specifically, the coefficients for voting (-.124) and political interest (.134) indicate that the belief that voting is a waste of time decreases the probability that immigrants will pursue citizenship, while being interested in politics increases the probability of naturalization. Thus, the statistical model shows that our expectation or hypothesis is correct, even after controlling for rival hypotheses. Our full sample model also shows that, controlling for other variables, the effect of being a woman (rather than a man) increases the likelihood of naturalization.

After observing the gender gap in naturalization, we became curious about whether the impact of political orientations on naturalization might be different for Latinas and Latinos. Qualitative research conducted by others[17] indicated that men and women experience migration and naturalization in the US differently. To systematically determine whether the influence of political orientations on naturalization is fundamentally different for men and women, we separated our initial sample and estimated the impact of political orientations separately for male and female Latino immigrants. Thus, we examined the impact of the independent variables on the probability of women undertaking naturalization, and the impact of the independent variables on the probability of men undertaking naturalization. The results in these models confirm our expectations – indicating that attitudes about voting significantly impact the choice to become citizens among women only, while interest in politics exerts a significant impact on pursuing citizenship for men but not women.

What do we make of these results? To understand our findings, we turn to the qualitative research on immigration and gender. Traditionally, research in this area[18] has found that women more often express interest in and knowledge about local politics, while men are more likely to express interest in politics at the national level. The results of our survey bear this out, with more Latinos expressing interest in politics than Latinas. For this reason, we are not surprised to find a weak relationship between political interest and citizenship among Latinas. Our results also demonstrate that, while both Latinos and Latinas are equally likely to disagree with the statement "voting is a waste of time," this attitude influences the choice to become citizens among Latinas only. Again, qualitative research[19] provides a guide for us. This research indicates that immigrant Latinas tend to be more active in community and local politics than Latinos. We suspect that participation in community and locally centered activities may lead women to naturalize and participate in US politics at a greater rate than their male counterparts.

Lessons to Be Learned

Historically, much of our knowledge of the Latino population was based on

anecdotal accounts which relied on cultural stereotypes to explain their behavior and attitudes. Social scientists who sought to challenge these ideas could only rely on in-depth interviews with a few respondents, since public opinion surveys with large Latino samples were unavailable. Despite their best efforts to present an alternative narrative, their research was often challenged, since the findings could not be generalized beyond the individuals interviewed. It was not until the 1989 *Latino National Political Survey* (LNPS) that social scientists could finally make general claims about the Latino population by having access to large *n* survey data. The LNPS allowed scholars to dispel many commonly held stereotypes about Latinos through the application of quantitative or statistical methodologies.

One of us (Adrian Pantoja) recalls, as an undergraduate, learning about the 1989 LNPS and being impressed with the fact that a handful of political scientists could shape public attitudes and policy through survey data and **quantitative analysis**. In this present research effort we continue in this tradition, and we hope that our research makes not only a theoretical contribution but a political once as well. We believe our research question meets the criteria of theoretical and political importance and encourage the readers of this chapter (perhaps future political scientists) to pursue research questions with the same goals in mind.

In our study we empirically demonstrated that Latinos with positive political orientations would pursue naturalization at higher rates relative to similarly situated Latinos lacking these orientations. Moreover, we found that Latinas (women) pursue naturalization at higher rates than Latinos (men) and that the reasons for this pursuit vary by gender. A skeptic could take the same data set we use, replicate our methods and reach the same conclusion. Despite our confidence in our findings, we do not want to suggest that the study is not without any flaws, or that any scholar knows definitively why Latinos choose to pursue naturalization. We have merely added one piece to the puzzle. But the puzzle is not solved. In the future, scholars must continue to examine this question through the use of different data, measurements and methodologies for us to fully understand the forces driving Latino immigrants to become US citizens. Understanding the key factors limiting or fostering the acquisition of citizenship among Latinos

could help unlock the door that allows Latino voters to achieve their full political potential.

Interested to Know More about the Study Discussed in this Chapter?

Consult the research publication:

> Pantoja, Adrian D., and Sarah Allen Gershon. 2006. "Political Orientations and Naturalization Among Latino and Latina Immigrants." *Social Science Quarterly* 87(5): 1171–1187.

Exercises and Discussion Questions:

1. What is the advantage of employing a large *n* or quantitative statistical approach?
2. What is the disadvantage of employing a large *n* or quantitative statistical approach?
3. Why did the authors conduct a multivariate statistical analysis?
4. Are the sample and population the same thing?
5. Why was the LNPS so important to experts in Latino politics?

Recommended Resources:

Pew Hispanic Center (www.pewhispanic.org/)
University of Washington Institute for the Study of Ethnicity, Race and

Sexuality (www.depts.washington.edu/uwiser/)

University of Washington Data Archive, Homepage of Professor Matt Barreto (www.faculty.washington.edu/mbarreto/data/index.html)

Inter-University Consortium for Politics and Social Research (www.icpsr.umich.edu/icpsrweb/ICPSR/)

Public Policy Institute of California (www.ppic.org/main/home.asp)

The Roper Center, Public Opinion Archives (www.ropercenter.uconn.edu/)

Notes

1 https://www.politico.com/magazine/story/2016/06/clintons-latino-firewall-213966

2 https://www.thenation.com/article/when-anti-immigrant-politics-came-back-to-haunt-the-republican-party/

3 Pantoja, Ramírez and Segura 2001: 729–750.

4 For a closer look at these polls, see www.latinodecisions.com.

5 http://www.pewhispanic.org/2017/09/18/facts-on-u-s-latinos/

6 www.harvardhispanic.org/what-will-it-take-to-awaken-the-sleeping-giant-latino-issues-in-the-2016-presidential-election/

7 Aron, Aron and Coups 2006.

8 Aron, Aron and Coups 2006.

9 Hardy-Fanta 1993; Jones-Correa 1998.

10 Quote taken from our article, Pantoja and Gershon 2006, 1172.

11 Pantoja and Gershon 2006: 1181.

12 Pantoja and Gershon 2006: 1181.

13 See Johnson and Reynolds 2008 for a deeper discussion of levels of measurement.

14 See Aron, Aron and Coups 2006 for a more detailed discussion of ANOVA tests.

15 See Borooah 2002 for a more detailed discussion of ordered logistic regression.

16 Long and Freese 2006.

17 Pedraza 1991; Hondagneu-Sotelo 1994; Menjivar 2000; Levitt 2001.

18 Hardy-Fanta 1993.

19 Hardy-Fanta 1993; Jones-Correa 1998.

References

Aron, Arthur, Elain N. Aron, and Elliot E. Coups. 2006. *Statistics for Psychology,* 4th Edition. New York: Pearson.

Borooah, Vani K. 2002. *Logit and Probit: Ordered and Multinomial Models.* California: Sage Publications.

Hardy-Fanta, Carol. 1993. *Latina Politics, Latino Politics: Gender, Culture, and Political Participation in Boston.* Philadelphia, PA: Temple University Press.

Hondagneu-Sotelo, Pierrette. 1994. *Gendered Transitions: Mexican Experiences of Immigration.* Los Angeles, CA: University of California Press.

Johnson, Janet B. and Reynolds, H. T. 2008. *Political Science Research Methods.* Washington, DC: CQ Press.

Jones-Correa, Michael. 1998. *Between Two Nations: The Political Predicament of Latinos in New York City.* Ithaca, NY: Cornell University Press.

Levitt, Peggy. 2001. *The Transnational Villagers.* Los Angeles, CA: University of California Press.

Long, Jay Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata.* College Station, Texas: Stata Press.

Menjivar, Cecilia. 2000. *Fragmented Ties: Salvadoran Immigrant Networks in America.* Los Angeles, CA: University of California Press.

Pantoja, Adrian D., Ricardo Ramírez, and Gary M. Segura. 2001. "Citizens by Choice, Voters by Necessity: Patterns in Political Mobilization by Naturalized Latinos." *Political Research Quarterly* 54(4): 729–750.

Pantoja, Adrian. 2016 (November 7). "Latino Voter Wall Blocks Trump's Path to White House." *Huffpost,* https://www.huffingtonpost.com/latinodecisions/latino-voter-wall-blocks_b_12842834.html (last access June 22, 2018).

Pantoja, Adrian D. and Sarah Allen Gershon. 2006. "Political Orientations and Naturalization among Latino and Latina Immigrants." *Social Science Quarterly* 87: 1171–1187.

Pedraza, Silvia. 1991. ''Women and Migration: The Social Consequences of Gender.'' *Annual Review of Sociology* 17: 303–325.

Survey Research

Religion and Electoral Behavior in the United States, 1936–2016

Lyman A. Kellstedt and James L. Guth

CONTENTS

Getting Curious: How does Religion Affect Voting?

When we and two colleagues sat down to write a book chapter titled "Faith Transformed," we had been studying religion and American politics for almost three decades, ignoring the ubiquitous maternal advice never to discuss the two in polite society. Although religion was largely ignored by political scientists when we did our graduate training in the 1960s and 1970s, we all had a personal interest in the subject. Two of us (Smidt and Green) were "preacher's kids," we were all active in Protestant churches, and we had all grown up in the Midwest, with its amazing smorgasbord of religious groups. Thus, we suspected that political scientists were missing something important by ignoring religion's impact on voting.

Real-world events finally stimulated scholarly attention to religion. The Christian Right's dramatic appearance in 1980, supporting the candidacy of Ronald Reagan, elicited a wave of studies by social scientists. We were well placed to take advantage of this new interest, as Kellstedt and Smidt were already exploring the role of evangelical voters, Guth had investigated the politics of Southern Baptist clergy, and Green was analyzing political contributors. Soon Smidt and Kellstedt were collaborating to explore religious voting, while Green and Guth studied religion among political activists and elites. By the late 1980s, we had "joined up" to study Protestant clergy, political activists and religious interest group donors.

The Research Strategy: What is Survey Research?

In the social sciences several research approaches are utilized to study political phenomena, including the collection and analysis of documentary evidence, ethnographic observation, experimental studies and elite interviews – all of which we have employed at times in our work. But most of our efforts involve **survey research**, the process of asking people questions, recording their answers and analyzing the results using statistical methods. Surveys can be conducted in several ways: by face-to-face interviews, telephone polls, self-administered mail questionnaires and, recently, by Internet contacts. Each has its advantages and disadvantages. Face-to-face interviews are valuable for establishing rapport and

often allow longer questionnaires, but are costly and rarely used today, with the exception of brief "exit polls" at voting places. Telephone interviews are much more common, but face increasing public resistance and new sampling problems arising from the pervasive use of cell phones. Self-administered mail questionnaires are employed most often to study members of organizations or activist groups. Internet polling is increasingly common, but still raises questions about sampling and representativeness.

Most of our early explorations involved mail surveys of activist groups and political elites, financed by small research grants, institutional funds, and often, our own pocketbooks. Our students helped by stuffing envelopes, entering data into computer files and doing preliminary analysis. But in studying religious voting, we were totally dependent on **secondary analysis** (analyzing data collected by someone else, perhaps for different purposes). Although a few specialized voting studies were available, the gold standard was the American National Election Study (ANES), begun with a small national survey in 1948 and continued in every presidential election since, as well as "off year" congressional elections beginning in 1958.

Nevertheless, we were quite frustrated with the way religion was measured in the ANES. At the very least, we wanted to describe how major religious groups behaved politically, but ANES questions were extremely crude. Before 1960, ANES asked only if a respondent was "Protestant, Catholic, Jew, Other, or None." That improved somewhat in 1960 when, prompted by the presidential candidacy of Catholic Senator John Kennedy, ANES produced a revised question which allowed us to differentiate members of Protestant religious families (Baptists, Lutherans, Methodists, Presbyterians, etc.), but not to identify many specific denominations within those families, denominations we knew had very different theological and political leanings.

Not until 1989 did the ANES devote one of its periodic **pilot studies** to developing new questions on religion. (Given the fierce competition for space on its surveys, the ANES uses pilot studies to develop new questions or modify old ones. They often use smaller samples than major surveys.) We worked with a committee drafting new religious items and subsequently analyzing them.[1] Kellstedt (joined later by Green) developed a new affiliation code which allowed us to assign the members of hundreds of American religious groups to

theoretically meaningful traditions. This task required detailed interviewer probes to produce a precise religious "belonging" measure, allowing us to differentiate, for example, between "Southern" and "American" Baptists or between members of the Lutheran Church, Missouri Synod (LCMS) and the Evangelical Lutheran Church in America (ELCA). This scheme was first employed in the 1990 ANES survey and, with minor modifications, has been used ever since. Better measurement of religious affiliation allowed us to classify voters into *religious traditions*, made up of denominations and local churches with similar beliefs and practices.

Religious traditions became our central concern in the early 1990s. Given the diversity of American religion and the small samples typical of social science research, we wanted to combine members of different denominations into meaningful categories, providing a large enough "N" (number of cases) for statistical analysis. Our religious tradition classification is historically rooted, designed to capture denominations and churches with similar theological tendencies and common organizational commitments. For us, the major American religious traditions are Catholicism, evangelical Protestantism, mainline Protestantism, Black Protestantism, Judaism and the unaffiliated or "seculars."[2] Smaller faiths must usually be arbitrarily relegated to an "other" category, but can be treated individually in very large samples, as we do later in this chapter. Most political scientists now use this classification, and many sociologists follow a similar one derived in part from our work.[3]

Getting affiliation right was important, but religion also includes belief and behavior components. Once again, we found ANES lacking. The only "belief" item was a crude three-option question on the Bible asked in 1964 and 1968, but dropped until 1980, when it reappeared permanently, along with a query on "born-again" status. In measuring religious behavior, ANES had always asked about church attendance and, in 1980, added items on the frequency of prayer and salience of religion. Although these questions were useful as far as they went, that was not very far, especially on religious belief. We had experimented with many belief and behavior items in surveys of college students, religious activists and political contributors in the 1980s and had worked with new measures in the 1989 ANES Pilot Study. Then in 1992 every social scientist's dream came true: the Pew Charitable Trusts, after supporting our earlier

exploratory work, agreed to fund our own large national survey.

The construction of the interview schedule proved a practical education in the intricacies of survey research. We confronted first the constraints of time and space. As we were conducting a telephone poll, it had to be relatively short (20 minutes or so) to maximize responses. Clearly, we had to make tough choices on question priority. We relied primarily on **closed-ended** items, providing response options, rather than **open-ended questions**, permitting the respondent to offer whatever came to mind. Thus, we asked respondents whether the United States should cut defense spending, and how strongly they agreed or disagreed, rather than, "What do you think about defense spending?" Although this risked forcing responses into a few categories and missing nuances of opinion, it also made coding answers much easier, and ensuring comparability. Wherever possible, we used "tried and true" questions already developed by respected polling organizations, in part to allow us to compare results with earlier surveys. If such questions did not exist, we wrote new ones, following the best practices in the discipline, avoiding loaded and complex questions, and emphasizing clarity in wording.[4]

Question order was also a matter of concern. We kept related questions in batteries of similar items, to reduce confusion on the part of the respondent, but at the same time kept some crucial items apart, so as to avoid undue mutual influence. And on political issues, we framed questions to reduce **response set**, where interviewees would simply "agree strongly" with all the statements (perhaps to get done with the survey). We also discovered that some questions were more sensitive than others and needed special placement. In our first pre-test, we found that starting the survey with religious questions produced a very high non-completion rate, as respondents simply refused to participate, or hung up soon after the survey began. When we reversed the order, putting the political questions first, the problem was solved. Once into the survey, respondents kept cooperating, even when the subject turned to religion.

Our goal was to develop **multi-item indices** of religious belief and behavior that were **reliable** (behaving the same way over time) and **valid** (showing expected relationships with other variables that should be related in theory). An **index** combines scores from several survey items designed to measure a single concept. For example, a religious behavior index might add responses on the

frequency of church attendance, scripture reading and personal prayer to provide a better measure of "religiosity." Such indices are vital for reducing the **measurement error** in single items. For example, a person might go to church each week under pressure from a spouse, but be uninvolved otherwise (never praying, giving money or singing in the choir). Combining answers to several questions will negate somewhat a "high" (or low) score on one item that does not reflect a "true" measure of overall religiosity. And, using an appropriate metaphor, we insist that "Measurement error is Sin."

After many fits and starts and hundreds of hours in front of our computers, we eventually produced indices of behavior and belief which were improvements over the ANES's reliance on single items. A five-item *religious behavior* index which included church attendance, prayer, financial contributions, involvement in religious small groups and scripture reading that worked across time (showing reliability), and had strong relationships with political behaviors like voting and working in political campaigns (demonstrating validity).

Developing a *religious belief* index proved more difficult. In constructing questions that went beyond the simplistic ANES Bible item, we confronted a basic issue: should we use "core" questions that could be answered meaningfully by people in almost all traditions (as well as by those with no religious faith)? Or should we design different batteries for Catholics and Protestants, Muslims and Jews? We went back and forth on this challenge. At one time or another all of us managed to be on each side of this issue, but we finally decided to emphasize common items, questions that elicited fundamental religious orientations from almost all Americans. This strategy kept surveys simple to administer (with no need for interviewers to "branch out" depending on the respondent's tradition), minimized the time required, allowed more space for other questions and facilitated data analysis.

After all this experimentation, by 2006 we had developed a five-item index based on queries about the existence of God, life after death, Holy Scripture, the Devil and evolution. Respondents had little trouble answering these questions and the index produced similar results over time (demonstrating reliability) and was meaningfully related to other religious variables and to vote choice, partisan preferences and policy attitudes (demonstrating validity). Now we had the conceptual tools to answer our research question: How had religious factors

influenced presidential voting from the New Deal era to the present?

Issues with this Research Strategy

As our discussion illustrates, survey research has both benefits and costs. Asking a representative sample of people questions about their affiliations, beliefs, opinions and behaviors is usually the best, and often the only way, to determine the distribution of these variables in the mass public or some part of it. Carefully conceived standardized questionnaires allow researchers to compare responses of different groups within the mass public, and open-ended questions can be used to explore topics of interest in more depth – and in the respondent's own words. Well-done survey research also permits scholars to trace public opinion and behavior over time, providing the material for **longitudinal analysis** of the sort done in this chapter. For scholarly work on public opinion and behavior, there is no substitute for modern survey research.

But there are important costs and limitations associated with survey research. As we discovered, secondary analysis of data gathered by others is often frustrating, as the original researchers may not have shared a later scholar's interest or expertise in a particular field of inquiry. This means that the questions of interest are sometimes rare, and often poorly conceptualized and measured. But the very significant burden of financing one's own national survey precludes most such efforts, unless underwritten by major foundations, government or other institutions. Indeed, the prohibitive expense of in-person interviews has produced alternative strategies of phone or Internet "interviewing," which still entail substantial costs and face various sampling problems. Finally, survey research requires substantial amounts of time, energy and attention for the production of valid and reliable measures of the concepts being investigated and for the efficient execution of the survey itself.

The Literature Review: Research on Religious

Alignments in American Politics

Producing better religious measures for national surveys was not just an exercise in curiosity, but would also allow us to test two competing views on religious influences on political behavior: the *ethnocultural* theory and the *restructuring* theory. We all had some training in history and were avid readers of the "ethnocultural historians" of the 1960s and 1970s, who confirmed our suspicions about the centrality of religious voting in earlier American history.[5] We were most impressed with Paul Kleppner's argument that nineteenth-century religious groups were divided between "pietists" (mostly Protestants), who supported first the Whigs and later the Republicans, and "liturgicals" (often Catholics), who backed the Democrats. Kleppner assigned myriad ethnoreligious groups to these categories and then painstakingly analyzed precinct voting and census records. (As he focused on our home region of the Midwest, his findings may have been especially convincing.) Unfortunately, ethnocultural historians did not venture far past 1896, leaving a yawning gap on ethnoreligious voting in the early twentieth century. Nevertheless, early social science studies of voting found continuing ethnoreligious divisions, with Northern Protestants supporting the GOP, and religious minorities, such as Catholics, Jews, Black Protestants and Southern evangelicals, forming the Democratic "New Deal coalition." Despite scholars' focus on the influence of social class on electoral choice after 1933, many Americans still voted on the basis of *ethnoreligious tradition*, as the 1960 election reminds us.

Although this historical literature "rang some bells" with us, we were also intrigued by a competing theory just emerging from the sociology of religion. In 1988, Robert Wuthnow argued that American religion had been "restructured." The divisions were no longer *between* but *within* religious traditions, engaging theological "conservatives" against "liberals." His argument was not about politics, but we saw the political implications. In 1991, James Davison Hunter's *Culture Wars* built on Wuthnow's thesis, positing *political* divisions between "orthodox" and "progressive" religious camps, especially on "hot button" social issues, such as abortion. We picked up on his suggestion in "It's the Culture Stupid!" (playing off a 1992 Clinton campaign motto, "It's the Economy, Stupid!"),

in analyzing the 1992 election.[6] We showed that there was some validity in the Wuthnow–Hunter *restructuring* or *culture wars* approach, as evangelical and mainline Protestants, as well as Anglo-Catholics, were indeed dividing as these authors predicted, but other groups' voting behavior still resembled that posited by the ethnocultural model. African-American Protestants, Jews, Latino Catholics and Latter-Day Saints (to name only the most obvious) still voted as cohesive ethnoreligious communities. We were also uncomfortable with Hunter's dichotomy of orthodox and progressives, knowing that any measure of theological "orthodoxy/progressivism" would have citizens arrayed along a continuum, not clumped in two groups. In religious warfare, as in politics, people often joined competing camps (think Republican and Democrat), but personal experience told us that many folks were "centrists" in these theological struggles. (Think of them as religious "Independents.") Thus, if we had to simplify theological factionalism within religious traditions, we preferred three categories to two, with *traditionalists, centrists* and *modernists* making up the major factions.

Conducting the Study: Collecting the Data and doing the Analysis

Thus, by 2006 we had developed the necessary concepts and measures to determine (1) how religious factors shaped presidential elections between 1936 and 2004 and (2) whether ethnoreligious politics had been replaced by culture war politics. But we still faced thorny obstacles to this project. The basic problem was simple: as we went back in time, religious measures available on national surveys became fewer and cruder. We wanted to confirm that ethnoreligious voting behavior had characterized the New Deal, but where could we find data for an era in which only one of us was alive, and not yet past bottle-sucking mode? We knew that Gallup began polling during the 1936 presidential election, and that these surveys were available at the Roper Center, a public opinion data archive. As we examined these polls, we had special requirements. We needed the most detailed religious affiliation measures possible (something better than

"Protestant, Catholic, Jew, Other, None"), and we hoped (and even prayed) for some belief and practice items (only a few were found). And, quite obviously, we required questions on presidential voting and partisanship. We found only a few surveys with good measures of both religion and political behavior, but enough to make informed estimates.

In **operationalizing** our idea of religious traditions (developing specific empirical measures for the concept), we faced some difficult choices. For example, Gallup grouped respondents by religious families (like "Lutheran"), but we wanted to assign them to religious traditions (such as evangelical or mainline Protestant). As most "Lutherans," "Presbyterians" and "Methodists" were in mainline denominations, we had to assign all of them to that category, despite sizeable numbers from smaller evangelical denominations within each "family." Similarly, all white "Baptists" had to be classified as evangelicals, as most were in evangelical denominations. Despite such limitations, the presidential voting data in Table 8.1 demonstrates that ethnoreligious voting was alive well into the 1940s, with mainline Protestants providing the strongest GOP presidential vote, and Roman Catholics and evangelicals anchoring the Democrats.

Here we should caution readers that Gallup surveys before 1948 were based on **quota sampling**. Gallup used the US Census to determine the populations of various demographic groups, such as young or old, Northern or Southern residents, black or white, *etc.* Interviewers were instructed to fill quotas from each group, based on their proportions in the population. After the fiasco in 1948 when Gallup incorrectly predicted that Thomas Dewey would defeat Harry Truman for the presidency, Gallup turned to **random sampling** procedures, in which all possible respondents throughout the country have an equal, or known, probability of being selected for interview. Although random sampling involves lots of practical problems, it has many advantages over quota sampling.

Random selection allows calculation of the **sampling error**, the difference between the sample estimate and the "real" **population** parameter that results from the fact that only a part of the population has been surveyed. For example, if the margin of sampling error is plus or minus 2 percent in a poll projecting that Obama will receive 52 percent of the vote, his "true" percentage could be as low as 50 or as high as 54. And random selection procedures allow one to talk about a **confidence interval**, the range of likely values for the "true" population value,

given the result from the sample for a given **confidence level** (for example, a 95 percent confidence interval means that the results are likely to have occurred by chance only 1 in 20 times). As quota samples are not equal probability samples, the laws of probability do not apply and margins of error and confidence intervals cannot be determined. Thus, results before 1948 must be evaluated with these limitations in mind.

Our use of ANES surveys from 1948 until 1984 presented fewer problems, but still required important conceptual and operational choices. Although we did not like ANES measurement of religious affiliation, we had to use it until we could employ our own surveys for 1988 (or use the new ANES measures available in 1990). Indeed, before 1960 the ANES had no breakdown of "Protestants" and, as a result, most scholarly work on religion and voting starts with 1960, when more detailed information was available. We hoped to do better. We discovered that ANES had conducted a **panel study** from 1956 to 1960 (panel studies interview the same people at several time points, while **cross-sectional surveys** interview different individuals at each). As we had the improved 1960 affiliation data on panel members, we could assume that affiliations did not change from 1956 to 1960—that a Southern Baptist in 1960 was a Southern Baptist in 1956. Although not universally true, the short-term stability of religious affiliation in that era made this a reasonable assumption. Even better, we found that the 1956 survey asked respondents to "recall" voting choices in both 1948 and 1952.

Thus, assuming continuity in religious affiliation and accuracy in recollection, we estimated religious group voting for both earlier years. Of course, we dropped panel members too young to vote in those elections, and we could not include those who voted in one or both years but died before 1956. Nevertheless, we see no persuasive reason why these problems should distort the findings dramatically. Despite some arbitrary assumptions, the results in [Table 8.1](#) comport with information about the 1948, 1952 and 1956 elections from other sources. In 1948 Black Protestants were moving dramatically toward the Democrats, reflecting both the Truman administration's appeal to Blacks and the States' Rights candidacy of Strom Thurmond. Both 1952 and 1956 are classic examples of "deviating" elections where many citizens ignored their party preference to vote for the opposition candidate. Folks did like Ike! One of us even voted for him. Note that the groups deviating the most—evangelical Protestants

and Anglo-Catholics—gravitated strongly toward the GOP later in the century, suggesting that they were already "on the move."

Did our painstaking efforts to obtain pre-1960 data pay off? Obviously, we think so. We may be engaging in wishful thinking, having invested enormous amounts of time and energy in the project, but if our assumptions are reasonable, we did show the ethnoreligious basis of the "New Deal coalition," something scholars can only speculate about if they begin analysis in 1960. Our findings were based on less than ideal data, but they came from large-scale *national* surveys and are consistent with local studies that documented religion's impact on voting.[7] For the years between 1960 and 1984, we used ANES data. The 1960 ANES affiliation codes were not perfect, but we were much more comfortable with them than with the "religious family" categories Gallup used for 1936–1944. And the measurement of religious tradition got better as ANES affiliation questions improved in 1990 and we could use our own surveys from 1988 until 2008. (We turned to the ANES again after our surveys ended in 2012 and also used the Cooperative Congressional Election Study.)

What did we find? As Table 8.1 demonstrates, in the years since the 1930s America's larger religious traditions "realigned" politically, with evangelicals becoming the strongest Republican voters, with 75 percent or more choosing the GOP candidate after 1996. Mainline Protestants and white Catholics left their respective GOP and Democratic "homes," becoming "swing" groups closely divided between the parties, while Black Protestants and secular voters became increasingly strong Democratic constituencies. Despite differences in

Table 8.1 Republican Percent of the Two-Party Vote for President by Major Religious Traditions, 1936-2012

Religious Tradition	1936	1940	1944	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000	2004	2008	2012	Gain/Loss
Evangelical Protestant	36	46	48	38	63	60	60	38	69	84	51	65	74	69	69	67	74	78	75	78	+42
Mainline Protestant	48	58	60	55	72	71	70	46	72	75	64	70	72	62	57	55	60	50	50	59	+11
Black Protestant	35	38	32	6	20	36	32	0	4	16	7	7	11	8	10	11	4	17	5	4	−31
White Catholic	18	28	33	25	49	55	17	22	40	64	44	58	55	51	46	46	50	53	51	53	+35
Unaffiliated	28	41	39	37	56	53	45	32	46	53	44	59	57	50	34	43	36	28	27	32	+1
Total Sample	36	45	48	41	58	60	51	33	54	64	49	56	58	53	47	47	50	51	46	48	+12

1936–1944 Gallup polls, AIPO 0149 Forms A & B, AIPO 0208, AIPO 0209, AIPO 0210, AIPO 0211, AIPO 0308, AIPO 0335, and AIPO 0360. For all religious groups in 1948 and for evangelical and mainline Protestants in 1952 and 1956, NES 1956–1960 Panel Study; 1952–1984 NES Cumulative File; 1988–2004 National Surveys of Religion and Politics, University of Akron: 1992, 1996, 2000, 2004, and 2008; NES 2012.

our data sources and their limitations, [Table 8.1](#) tells a crucial story about trends in ethnoreligious voting in presidential elections.

Another way to look at these patterns is by "presidential vote coalitions." [Table 8.2](#) reports the religious composition of the two-party vote for selected elections since 1936. In the earliest years, the GOP was clearly a party of white Protestants, especially mainliners, while Democrats drew significant backing from mainliners, most Catholics and many evangelicals, as well as representation from smaller groups. By the time of the 2000 election, dramatic changes had occurred: evangelicals provided the largest bloc of Republican votes, with mainline Protestants and white Catholics each supplying around one-fifth of the GOP total. The Democratic religious constituency became increasingly diverse, but the unaffiliated emerged as the largest single group, followed by roughly equal numbers of mainliners, Black Protestants, white Catholics and those from "all other" religions. [Table 8.2](#) also shows the gains and losses of each religious group as a component of each party coalition over the period. As parties do respond to their voting constituencies, these changes had profound implications for public policy.

We have traced developments in ethnoreligious politics, but to test whether such alignments were giving way to culture war divisions, we had to use ANES data for the 1970s and 1980s, the critical period for this transition. We combined the ANES's one belief measure (the Bible item) and one behavior measure (church attendance) in a crude "religious traditionalism" classification.[8] The results confirmed our expectation that the ethnoreligious model still defined voting patterns in the 1960s. By the 1980s, however, evangelicals fit a culture wars model, as traditionalists were more Republican than their centrist and modernist co-religionists.[9] Given the emergence of the Christian Right, based in evangelicalism, this result comported nicely with events in the political world.

With the 1988 election, we could use recall data from our 1992 National Survey of Religion and Politics (NSRP) to assess more fully the emergence of a modified culture wars pattern. Using multi-item religious belief and behavior indices, we had a more sensitive test of religious restructuring. In addition, NSRP samples were much larger than those in ANES, increasing confidence in the findings. For example, ANES data suggested that the culture-wars model had kicked in by 2000–2008 for white evangelicals, mainliners and Catholics, with differences

between traditionalists and modernists ranging from 10 to 18 percentage points. For the same period, our NSRP data in Table 8.3 shows much larger gaps, ranging from 22 percentage points among evangelicals to 36 among white Catholics during this period. By 2012, even the ANES data strongly confirmed this development, despite less than adequate measures for traditionalism. Indeed, the culture wars theory explains the close party balance among mainline Protestants and white Catholics, as these traditions are the most theologically factionalized. All this shows the importance of using reliable and valid indices of religious beliefs and behaviors. It may seem a bit complex, but the results warranted the effort.

Table 8.2 Presidential Vote Coalitions by Party and Religious Traditions, 1936-2012 (percent of total party vote)

Religious Tradition	1936		1960		1984		1992		2000		2004		2008		2012		Gain/Loss	
	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D
Evangelical Protestant	19	21	24	17	24	12	37	15	38	13	40	12	39	11	33	8	+14	−13
Mainline Protestant	59	37	58	25	30	16	23	16	22	15	18	19	20	18	18	11	−41	−26
Black Protestant	1	1	3	6	2	18	2	14	1	18	3	13	1	16	1	17	0	+16
White Catholic	9	24	7	36	19	22	23	24	21	20	20	19	23	20	19	15	+10	−9
All Others	5	8	3	10	7	13	7	18	8	17	11	15	10	15	14	22	+9	+24
Unaffiliated	7	9	5	6	18	19	8	13	10	17	8	22	7	20	15	27	+8	+18
Total Sample	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100		

Legend: R = Republican; D = Democratic
Sources: see Table 8.1

A final important question is whether these divisions are long-term ones, or an artifact of the politics of the 1980s and 1990s. Perhaps the extraordinary 2016 electoral victory of Donald Trump, a "populist" candidate with only vague religious ties, might be expected to have altered previous alignments. And the changing nature of American religion, with its greater diversity, declining numbers of white Protestants, and increasing numbers of unaffiliated voters might also result in important changes.[10]

What did religious alignments look like in 2016? To answer this question, we use the 2016 Cooperative Congressional Election Study (CCES). This survey has good affiliation measures and a gigantic sample (*N*=64,593), allowing us to examine the political behavior of many smaller religious groups and different kinds of "unaffiliated" voters with some confidence. Unfortunately, the CCES has no religious belief items, but does have questions on the importance of religion,

religious service attendance and personal prayer that we can use to compute a traditionalism/modernism index.[11]

In Table 8.4 we find that the 2016 patterns among religious groups remained broadly familiar. Trump actually received a larger proportion of the evangelical Protestant vote than any GOP presidential candidate since 1972 and matched the typical recent GOP nominee among mainline Protestants and white Roman Catholics. And within all three large white Christian traditions, the culture war differences are also quite familiar, with traditionalists most Republican, followed by centrists, then modernists. (Note that Trump did better than most recent GOP candidates among the modernists.) Given the huge CCES sample, we can also have considerable confidence in our estimate for smaller religious groups. As in the past, Latter-Day Saints were mostly Republican, although less so than in 2012 when fellow Mormon Mitt Romney was the nominee; Mormons were joined in their GOP preferences by Eastern Orthodox voters.

As we expected, Black Protestants remained solidly in the Democratic camp. Interestingly, Latino voters were divided by religious affiliation: evangelicals gave Trump a solid majority, but Latino mainliners and Catholics favored Hillary Clinton by large margins. Muslims, Buddhists and Hindus were also overwhelmingly Democratic, while our residual category of "other religions" also leaned solidly to the Democratic side. Thus, the Democratic Party maintained its historic position as the home of most ethnoreligious minorities. In a related vein, those claiming no religious preference whatever tended to vote Democratic, with self-described atheists almost unanimous in that choice, followed by agnostics, and at a distance by those who reported their religion as "nothing in particular."[12] The strength of Republican and Democratic partisanship within each group tells much the same story, although the number identifying with the parties is naturally somewhat lower than those voting for their candidates, as Independents and others are omitted.

When we turn to the religious basis of the party voting coalitions (compare columns 4 and 5 with Table 8.2), only a little seems to have changed since 2000. The GOP coalition is still overwhelmingly white Christian, with evangelical and mainline Protestants, Anglo-Catholics and Latter-Day Saints constituting about 70 percent of Trump voters. In contrast, the Democratic coalition is loaded with ethnoreligious minorities – Black Protestants, Latinos, Jews – and many

unaffiliated or secular folks (who now constitute fully one-third of the Democratic electorate). On the culture war side, traditionalists still support the GOP in large numbers, while modernists back the Democrats. Thus, the religious party coalitions in 2016 resemble those going back at least to 2000, although the declining numbers of white evangelicals has reduced their dominance in the GOP coalition, and the growth of the "unaffiliateds" has had the opposite effect in the Democratic voting bloc (cf. Table 8.3).

We labored mightily on this chapter, and we leave it up to our readers to assess the results. From our perspective, we have demonstrated that the old ethnoreligious electoral patterns have been substantially altered since the 1930s. Although many religious minorities still display those propensities, favoring the Democrats – the traditional home of religious minorities – among evangelical Protestants, mainline Protestants and Anglo-Catholics internal religious divisions result in very different political choices, with traditionalists favoring the Republicans and modernists backing the Democrats (as do the growing numbers of unaffiliated voters). These changes raise all sorts of further questions: Why did they occur and what are the policy implications? These are critical topics for further investigation.[13]

Table 8.3 Republican Vote for President for Major Religious Traditions, Controlling for Traditionalism, 1988-2012 (in percent)

Religious Tradition and Group	1988	1992	1996	2000	2004	2008	2012	Gain/Loss
Evangelical Protestant	69	69	67	74	78	76	78	+9
Traditionalist	74	83	81	87	88	88	86	+12
Centrist	65	68	58	63	70	68	73	+8
Modernist	64	44	43	43	56	54	36	−28
Traditionalist/Modernist Gap	**+10**	**+39**	**+38**	**+44**	**+32**	**+34**	**+50**	**+40**
Mainline Protestant	62	57	54	60	50	50	59	−3
Traditionalist	63	64	65	76	65	68	68	+5
Centrist	61	59	53	52	49	49	63	+2
Modernist	63	50	46	54	39	39	30	−33
Traditionalist/Modernist Gap	**0**	**+14**	**+19**	**+22**	**+26**	**+29**	**+38**	**+38**
White Catholic	51	46	46	50	53	51	53	+2
Traditionalist	51	54	54	61	74	58	75	+24
Centrist	48	47	49	49	52	63	53	+5
Modernist	54	39	37	39	35	34	24	−30
Traditionalist/Modernist Gap	**−3**	**+15**	**+17**	**+22**	**+39**	**+24**	**+51**	**+54**
Total Sample	**53**	**47**	**47**	**49**	**51**	**46**	**48**	**−5**

Sources: National Surveys of Religion and Politics, University of Akron, 1988–2008; National Election Studies 2012.

Religion and the 2016 Presidential Vote (in percent)

	Trump Vote	Republican Party ID	Democratic Party ID	GOP Vote Coalition	Dem Vote Coalition
Major Religious Traditions:					
Evangelical Protestant	**82**	**66**	**19**	**32.4**	**6.9**
Traditionalist	*86*	*71*	*15*	*22.4*	*3.4*
Centrist	*74*	*57*	*25*	*8.6*	*2.8*
Modernist	*67*	*52*	*28*	*1.4*	*0.7*
Traditionalist/Modernist Gap	*+19*	*+19*	*–13*	*--*	*--*
Mainline Protestant	**59**	**48**	**36**	**16.5**	**11.1**
Traditionalist	*65*	*54*	*33*	*6.6*	*3.4*
Centrist	*58*	*48*	*35*	*7.4*	*5.1*
Modernist	*48*	*39*	*42*	*2.5*	*2.6*
Traditionalist/Modernist Gap	*+17*	*+15*	*–9*	*--*	*--*
Anglo-Catholic	**58**	**45**	**39**	**19.1**	**13.2**
Traditionalist	*66*	*52*	*35*	*8.4*	*4.1*
Centrist	*55*	*41*	*41*	*7.8*	*6.1*
Modernist	*48*	*38*	*44*	*2.9*	*3.0*
Traditionalist/Modernist Gap	*+18*	*+14*	*–9*	*--*	*--*
Black Protestant	**8**	**6**	**83**	**1.0**	**11.5**
Smaller Religious Traditions:					
Latter-Day Saints	**70**	**63**	**19**	**1.7**	**0.7**
Eastern Orthodox	**63**	**45**	**39**	**0.7**	**0.4**
Latino Evangelical Protestants	**54**	**35**	**40**	**1.1**	**0.9**
Latino Mainline Protestants	**30**	**26**	**60**	**0.1**	**0.3**
Jewish	**28**	**27**	**61**	**1.5**	**3.8**
Latino Catholic	**25**	**18**	**65**	**1.5**	**4.1**
Buddhists	**22**	**20**	**60**	**0.3**	**1.1**
Hindus	**22**	**12**	**62**	**0.1**	**0.3**
Muslims	**12**	**9**	**76**	**0.1**	**0.8**
All other religions	**44**	**28**	**45**	**9.3**	**11.3**
Religiously Unaffiliated:	**29**	**19**	**53**	**14.6**	**33.6**
Nothing in Particular	*38*	*22*	*45*	*10.2*	*15.7*
Agnostic	*24*	*18*	*62*	*2.8*	*8.3*
Atheist	*14*	*13*	*68*	*1.6*	*9.6*
Total Sample	**48.8**	**35**	**45**	**100**	**100**

Source: Cooperative Congressional Election Study 2016.

Lessons to Be Learned

What lessons about research have we learned? First, we suggest choosing projects that interest you and draw on your own experience. Personal interest means you will stick with the enterprise even when the work is not going well, and personal experience is often a vital source of insight. At the

same time, there is no substitute for careful scholarly preparation, whether for a long-term investigation or a class project. We drew both inspiration and insight from our own experience and from the scholarship of others, not only in political science, but in history and sociology as well. Remember that the world is not really apportioned by the artificial divisions of the academy: you can garner important ideas from many sources. You also learn a lot by collaborative projects; it is safe to say that this chapter would never have been written had we not spent years exchanging and testing each other's ideas – and those of other scholars with whom we worked.

Another lesson is the importance of paying close attention to the data, whether you are using someone else's survey or constructing your own. First, know about the quality of the surveys. Were the samples chosen by random selection procedures? What were the sample sizes? The response rates? Second, how adequate were the questionnaires? Do they include measures of all the relevant variables? Do these questions really get at the concepts they are supposedly tapping? Are the questions clear or ambiguous? In the modern world, we are inundated by data drawn from polls, whether in newspapers, TV, magazines or, for that matter, the classroom. We all need to approach such data with a critical and informed eye.

The third lesson is one we learned originally in graduate school: multiple measures of key concepts (like religious beliefs and behaviors) are almost always better than one. Although limits on survey space often lead scholars to rely on single items, this decision usually comes at the cost of greater measurement error. Of course, any item can be worded in ambiguous fashion, and, as a result, each must be examined carefully for reliability and validity, whether it is to be used alone or as part of a composite measure.

A final lesson from our experience is that it takes a long time for scholars in any discipline to adopt "best practices" in addressing any intellectual issue. We have worked in this field for almost 30 years and have learned a lot, abandoning many "false leads" we once followed. Yet we often see younger scholars or those new to the study of religion and politics repeating the same mistakes we made long ago. Eventually, most will learn through trial and error, although they could shorten that period with careful

examination of previous scholarship. Of course, there are always those who draw from different theoretical traditions, or who obstinately insist on seeing the world of religion and politics in a different way than we do. They are wrong, of course, but we can still learn a lot from them!

Although years of work went into this chapter, we also had an enormous amount of fun. There is an old saw about the "loneliness of the scholar." We never felt that sensation: we enjoyed endless hours thinking, arguing and computing together (with more than a few social interludes). And we are part of a larger (and largely congenial) intellectual community engaged on a common quest to understand the nexus between religion and politics and to convey what we have learned to students, scholars, journalists and the general public. It really doesn't get much better than that.

Interested to Know More about the Study Discussed in this Chapter?

Consult the publication:

Kellstedt, Lyman, John Green, Corwin Smidt, and James Guth. 2007. "Faith Transformed: Religion and American Politics from Franklin D. Roosevelt to George W. Bush." In *Religion and American Politics*, 2nd edn., ed. Mark A. Noll and Luke E. Harlow. Oxford: Oxford University Press.
The authors thank Professor Noll for his gracious permission to reproduce modified versions of three tables from the original text.

Exercises and Discussion Questions:

1. In your assessment, what are the major strengths and weaknesses of

survey research?

2. How would you attempt to alleviate or minimize the weaknesses of such research?

3. Are any topics too "sensitive" to be studied by survey research? If so, what are they? Are there ways you might minimize this problem?

4. This chapter focuses on the impact of religion on vote choices. Are the substantive results of the research project discussed convincing to you? Why or why not? How could the researchers have done a better job?

5. Would you be willing to be interviewed if called on the phone? Why or why not? What does this tell you about contemporary obstacles to survey research?

Recommended Resources:

American Association of Public Opinion Research (www.aapor.org/): Sets standards for the conduct of survey research.

National Polling Organizations:

Gallup Polls (www.gallup.com/)
Pew Forum on Religion and Public Life (www.pewforum.org/)
Public Religion Research Institute (www.prri.org/): Regular surveys on American religion.
Pew Research Center for People and the Press (www.people-press.org/): Regular surveys of the American public on a vast array of issues. Data usually available for secondary analysis six months after survey date.
Mitofsky International (www.mitofskyinternational.com/): election exit polls

Social Science Data Archives:

General Social Surveys (www.norc.uchicago.edu/GSS+website/)

Inter-University Consortium for Political and Social Research (www.icpsr.umich.edu/icpsrweb/): The American National Election Studies time series and many other data sets are available here.

The Association of Religion Data Archives (www.thearda.com/): Many important studies of religion are available here.

Selected Scholarly Studies of Religion and American Politics:

Campbell, David, ed. 2008. *A Matter of Faith.* Washington, DC: The Brookings Institution.

Layman, Geoffrey. 2001. *The Great Divide.* New York: Columbia University Press.

Putnam, Robert, and David Campbell. 2010. *American Grace.* New York: Simon & Schuster.

Smidt, Corwin, Lyman Kellstedt, and James Guth, eds. 2009. *The Oxford Handbook of Religion and American Politics.* New York: Oxford University Press.

Notes

1 See Leege and Kellstedt 1993 for discussion of the 1989 Pilot Study and ANES religious measures derived from that experiment.

2 The most important issues involve differentiating evangelical Protestants, more theologically conservative, from mainline Protestants, who tend to be more liberal. As "denominational families" are often split between these traditions, deep knowledge of religious history and theological leanings is required for accurate classification. See Smidt, Kellstedt, and Guth 2009.

3 See Steensland *et al.* 2000.

4 A classic and still valuable guidebook for questionnaire construction is Converse and Presser 1986.

5 For a summary of the literature we read, see McCormick 1974.

6 Kellstedt *et al.* 1994.

7 See Lazarsfeld, Berelson, and Gaudet 1948; Berelson, Lazarsfeld, and McPhee 1954; Lenski 1963.

8 As the Bible item was asked in 1964 and 1968 and only picked up again in 1980, we cannot do this analysis for 1972–76.

9 Traditionalists were those who said the Bible was literally true and attended church frequently. Modernists were those who believed the Bible was the work of man, and not God, and was of little value, while rarely or never attending church. Centrists fall between traditionalists and modernists on one or both questions. For the full results from this analysis of ANES data, see Kellstedt *et al.* "Faith Transformed," 288–289.

10 See Cox and Jones 2017.

11 Although the CCES survey has many advantages, the administration of the survey permitted many respondents to use follow-up probes on religious affiliation that were not appropriate for their initial answers. Thus, the coding of respondents into religious traditions requires careful review of the available information on each respondent. A full review may produce slightly different results than reported here.

12 The huge sample also allowed us to estimate the Republican vote among even smaller religious groups than those reported in Table 8.4; for example, Asian Protestants (46 percent), Asian Catholics (34 percent), Wiccans (22 percent), "spiritual but not religious" (22 percent); and Jehovah's Witnesses (11 percent).

13 For a start, see Kellstedt and Guth 2014.

References

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting.* Chicago: University of Chicago Press.

Converse, Jean M., and Stanley Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire.* Beverly Hills, CA: Sage.

Cox, Daniel, and Robert P. Jones. 2017. "Additional Evidence of White Evangelical Decline." https://www.prri.org/research/american-religious-

landscape-christian-religiously-unaffiliated/

Kellstedt, Lyman A., and James L. Guth. 2014. "Religious Groups as a Polarizing Force." In *Polarized Politics: The Impact of Divisiveness in the US Party System*, ed. William Crotty. Boulder, CO: Lynne Rienner.

Kellstedt, Lyman, John Green, Corwin Smidt, and James Guth. 2007. "Faith Transformed: Religion and American Politics from Franklin D. Roosevelt to George W. Bush." In *Religion and American Politics*, 2nd edn., ed. Mark A. Noll and Luke E. Harlow. Oxford: Oxford University Press.

Kellstedt, Lyman A., John C. Green, James L. Guth, and Corwin E. Smidt. 1994. "Religion and the 1992 Election: It's the Culture, Stupid!" *First Things* 42(April): 28–33.

Lazarsfeld, Paul F., Bernard R. Berelson, and Helen Gaudet. 1948. *The People's Choice*. New York: Columbia University Press.

Leege, David C. and Lyman A. Kellstedt, ed. 1993. *Rediscovering the Religious Factor in American Politics*. Armonk, NY: M.E. Sharpe.

Lenski, Gerhard. 1963. *The Religious Factor*. Garden City, NY: Doubleday-Anchor.

McCormick, Richard. 1974. "EthnoCultural Interpretations of Nineteenth-Century American Voting Behavior." *Political Science Quarterly* 89(2): 351–377.

Smidt, Corwin E., Lyman A. Kellstedt, and James L. Guth. 2009. "The Role of Religion in American Politics." In *The Oxford Handbook of Religion and American Politics*. ed. Smidt, Kellstedt, and Guth. New York and Oxford: Oxford University Press.

Steensland, Brian, Jerry Park, Mark Regnerus, Lynn Robinson, W. Bradford Wilcox, and Robert Woodberry. 2000. "The Measure of American Religion." *Social Forces* 79(1): 291–318.

CHAPTER 9

Public Policy and Program Evaluation

Does High School Type Affect College Success?

David J. Fleming, Joshua M. Cowen and Deven Carlson

CONTENTS

Getting Curious: Evaluating Public Policy

Think back to your high school days. The awkward dances. The Friday night football games. The stress of completing college applications. Do you ever wonder how your life today would be different if you had attended a different high school? Would you be at the same university that you are now, or even at a university at all? How sure are you that you would have even graduated from high school? Our experiences as teenagers can have long-lasting impacts on our lives. These types of topics aren't just what daydreams are made of. For decades, public policy researchers have tried to examine how the choice of middle or high school can affect student outcomes. These analyses often compare how the performance of public school students differs from that of private school students.[1] Some people presume that private schools are more effective on average than public schools are, but is that really true? Answering that question may be more difficult than it sounds.

If a simple comparison between private and public high school students indicates that private school students have higher test scores or are more likely to attend college, we still can't be sure that private schools are more effective. This is because of **selection bias**. Parents and students who can afford to attend private schools and choose to do so may be fundamentally different from parents and students in public schools. Therefore, raw differences in student outcomes between private and public school students may not indicate that one sector is more effective than the other; rather, these differences may just reflect the type of parents and students who select into private schools and into the public sector.

Given these fundamental challenges in examining the efficacy of private versus public schools, how can we answer this question? Scholars, including us, are evaluating school voucher programs to learn more about how attending private schools may affect student outcomes. Voucher programs provide publicly funded scholarships to parents that they can use for private school tuition for their children. These programs provide access to private schools for many students who could not afford private school tuition and would otherwise attend public schools. The politics surrounding the school voucher debate is heated for a number of reasons. For example, questions of the separation of church and state arise. Further, where students attend school has been an issue linked to both desegregation and to segregation in the past. Teachers' unions and pro-market interest groups donate millions of dollars to shape public opinion and the actions

of elected officials on the issue of school vouchers.

Do school vouchers improve student performance and empower parents? Do voucher programs drain support from public schools and lead to a more unequal educational system? As interest groups, public officials and commentators make claims about the effects of school voucher plans, one wonders whom to believe. The presumptions and arguments made by advocates on both sides of the voucher debate need to be verified through a careful empirical study. This is particularly true as the number of school voucher programs continues to increase. Currently, there are 26 school voucher programs across 15 states and the District of Columbia.[2],[3]

Political scientists have informed and helped to shape the school voucher debate. Official evaluations of publicly funded voucher programs in Milwaukee, WI; Washington, DC; and Louisiana have been led by political scientists.[4] In many ways, this body of research began in Milwaukee, where University of Wisconsin political scientist John Witte served from 1990 to 1995 as the official state evaluator of the nation's first large-scale, publicly funded voucher program, the Milwaukee Parental Choice Program (MPCP). During the first year of Witte's evaluation, the MPCP had 341 voucher students in seven secular private schools. After his evaluation ended in 1995, the state of Wisconsin expanded the voucher program to include religious schools in Milwaukee. As of the 2016–2017 school year, the Milwaukee voucher program enrolled almost 28,000 students in 121 different religious and secular private schools.[5] This makes the MPCP the largest urban voucher program in the United States.[6]

During this period of tremendous growth in the MPCP, there were no scholarly evaluations of the program from 1996 to 2006 (when our evaluation began). The citizens of Wisconsin had very little information on how their tax dollars were being spent. Elected officials in the state were unsure about the possible positive or negative effects of the program. Scholars, activists and policymakers from around the country wanted to know if this experience with school vouchers was working to ameliorate the inequities and low performance that many identify in our urban school systems.

In 2006, the Wisconsin legislature enacted Act 125. Amongst other changes to state law, the legislation instituted a new standardized testing system and evaluation program for voucher students. All voucher schools were required to

administer standardized tests to students and submit the test scores to the School Choice Demonstration Project (SCDP), a nationally recognized research center based at the University of Arkansas and specializing in evaluating school choice programs. Further, the state of Wisconsin directed the SCDP to conduct a five-year study comparing voucher student performance to similar students in public schools. As three of the researchers working with the SCDP during this time, this meant that we now had the opportunity to examine the effects of the MPCP. Our work began in the fall of 2006, and the last round of data collection officially ended in the fall of 2010.

In this chapter, we discuss the results from one dimension of our multi-faceted study. We were interested in how experience with private schooling through school vouchers affects student attainment. Student attainment means reaching a given level of schooling. Specifically, we decided to examine how participating in the voucher program affected the likelihood that students would graduate high school in four or five years, the percentage that enroll in a two-or four-year university, and persistence, meaning a student remains enrolled in a university after their first year in college. These were the **dependent variables**, or outcomes, that we examined in our study.

This research originated as a **program evaluation**, a type of study with a long tradition in political science. Program evaluation is "the application of empirical social science research methods to the process of judging the effectiveness of public policies, programs, or projects."[7] These social science methods can involve many of the topics discussed in this book, including content analysis, survey research and statistical analysis. However, program evaluation is characterized less by the particular research design and more by the overarching goal to which the work is directed. Evaluators are interested in making causal statements about whether the program of interest "works" as its implementers intended. For example, do abstinence-only education programs decrease teenage pregnancy?[8] Or, has the Affordable Care Act increased Americans' access to healthcare?[9] Or, does Project DARE (Drug Abuse Resistance Education), a school-based drug prevention program, effectively reduce the number of students who use drugs?[10] Program evaluation is the approach used to answer these questions.

The Research Strategy: Quasi-Experimental Panel Study using Data Analysis

Our research strategy for this study, a quasi-experimental panel study using data analysis, is a mouthful, but each part of the design is important. The first component is the **quasi-experimental** design.[11] Researchers use quasi-experimental designs when they are interested in estimating a causal effect of a variable on an outcome, but are unable to use random assignment to treatment and control conditions, like in a true experiment. Previous evaluations of voucher programs have used a type of experiment, called a **field experiment**, or a **randomized control trial** (see Chapter 11).[12] Researchers may use this approach when there are more students applying for a voucher than there are spaces available. A random lottery then decides who will receive a voucher. This means that the lottery winners and the lottery losers will be very similar. Any differences in student outcomes would be due to winning or losing the voucher lottery. This is typically the preferred technique to deal with selection bias. However, in our study, we did not have a situation in which there were more students applying for a voucher than spaces available. Therefore, we used a quasi-experimental approach that we detail below.

Since we are examining students over time, we needed to employ a **longitudinal** design. We used a **panel study** because we wanted to examine the *same* sample of students at different time points. This meant that we tracked the students who began our study in the 8th and 9th grades until they could be first-year or second-year college students.

Our research strategy combined two approaches to data analysis: primary source analysis and secondary source analysis. When using **primary source analysis**, the researcher directly collects the data and observes the phenomena. Political scientists use this approach when a researcher interviews a Member of Congress, conducts a survey or observes a protest. Alternatively, **secondary source analysis** uses indirect sources of data, as they are secondhand. Social scientists often rely on intermediaries to collect the data. This type of analysis uses unobtrusive methods of data collection and is sometimes called document analysis.[13]

Secondary data can come from a variety of sources including autobiographies, photographs, newspaper articles, databases and reports. Political science researchers can analyze secondary data quantitatively or qualitatively. One type of secondary source analysis is **content analysis**, which researchers can use to analyze different forms of communication (see [Chapter 10](#)). When using content analysis, researchers create a coding scheme that helps them to classify the media content. For example, researchers at the Wesleyan Media Project track political advertisements during election campaigns and code the different images, issues and themes of each television advertisement.

Published data are another form of secondary data. Published data are generally publicly available and have been collected by a government entity, private organization or other researchers. For example, the *Statistical Abstract of the United States* from the U.S. Census Bureau provides a host of statistics that can be used by researchers. The *Statistical Abstract of the United States*, which is published annually, is an example of a **running record**. Running records are ongoing, recurring and often uniform sources of data. **Episodic records** are unsystematic sources of document data, such as diaries or papers from presidential administrations. Public opinion polls conducted by Gallup or other organizations, Congressional roll call votes, the CIA's *World Factbook* and the *CQ Almanac* are among the many popular sources of published data. Data archives are another popular source of published data. The ICPSR data archive through the University of Michigan and the National Opinion Research Center at the University of Chicago has publicly available data that can be used to answer many political science research questions. It is quite common for political science researchers to combine various sources of published data in their research projects.

Given that the goal of our school voucher evaluation was to examine differences in student attainment, we used both primary and secondary source analysis. It would have been very difficult for the research team to get complete information on the graduation and college enrollment of all of the students in the study ourselves. Rather, we primarily relied on the graduation lists kept by the public and private high schools, as well as the National Student Clearinghouse of College Enrollment, which collects data on the vast majority of students attending two-and four-year universities in the United States. In addition to these

sources, we also received public student test scores and student demographic information from the Milwaukee Public School District, and these data can be considered published data.

However, as is often the case with secondary data, we were limited by what was available to us. For example, we wanted more information about why students may not have graduated and what their home lives were like. Therefore, we completed our own phone survey of the parents of study participants, which is an example of primary data. By combining primary and secondary data sources, we were able to fully implement our research strategy.

Issues with this Research Strategy

There is no one perfect recipe or research strategy in political science. Rather, researchers face trade-offs when selecting one research approach over another. That lesson is apparent in the three components of our study.

Our main research goal was to examine how private schooling affects the likelihood a student graduates from high school and attends college. One possibility would be to compare the graduation rates of students in public schools to those in the voucher program. Although this method sounds reasonable, it may suffer from **selection bias**. Selection bias occurs when individuals self-select into the groups you will study in your analysis, for reasons that are also related to your outcome of interest.

Let's examine an example of the problem of selection bias. Imagine an undergraduate student, Mary Justice, who is interested in attending law school. She knows that she will need to take the Law School Admission Test (LSAT), and she will need a good LSAT score to attend her preferred law school. Ms. Justice is considering taking an LSAT prep class, in which she will learn about possible test topics and test-taking strategies, but she is of modest means. Should she take the class? To answer the question, she asks her friends who have already taken the LSAT if they took the prep class. She discovers that those who took the LSAT class scored *higher* than her friends who did not take the class.

Did taking the LSAT class *cause* students' scores to increase? Students were

not randomly assigned to take the class or not take the class. Rather, students could choose to take the class. Examining her list of friends, Ms. Justice realizes that many of the students who decided to take the LSAT class were her most studious and disciplined friends. On the other hand, she noticed that those friends who were more likely to spend Friday night at a party rather than in the library did not take the LSAT prep class. Therefore, Ms. Justice was unable to determine if the LSAT class actually increased students' scores, or if selection bias had influenced her results.[14]

There are two strategies to deal with selection bias. One possibility is to perform an experiment, as we described above. We used a second method to deal with selection bias, a quasi-experimental, **controlled comparison design**. When using this approach, political scientists try to mirror an experiment by making the groups used in the analysis as similar as possible. In a quantitative study like ours, this is often done using **control variables** at the analysis stage. Control variables are variables that are held constant during an analysis in order to isolate the relationship between the variables of interest. A controlled comparison design can also be implemented at the sampling stage. Here, the goal is to create two samples that are alike in many ways, but different in terms of one important factor. This can be done in a small-n comparison case study approach (see Chapter 3) or in a large-n quantitative study like ours. We decided to use a **matching** procedure to account for selection bias. For every voucher student in our sample, we selected a public school student who was as similar as possible (keep reading to find out what we mean by "similar"). Therefore, any difference that we see between voucher and public high school graduation rates should be because of the type of school the students attended, and not because of the different types of students in each sector.

Panel studies require that the same sample be followed over time. There are additional challenges associated with implementing a panel longitudinal design. Perhaps the most important for our study was panel **attrition**. This can occur when individuals in a panel drop out of the study because they no longer want to participate, moved away, or for some other reason. As individuals change home addresses and phone numbers, it can be very difficult to track study participants. This problem gets worse the longer the panel study lasts. This is problematic if those who leave the study are different from those who remain. For example, if

the low-performing students in the voucher program drop out of the study, it may incorrectly appear as if the voucher program is having a positive effect on student attainment. Rather, the higher graduation rates of the voucher students may be because the low-achieving voucher students are no longer in the study. Panel attrition is another example of selection bias. Researchers can try to minimize panel attrition by providing incentives to subjects to keep them participating in the study.

As you can see, quasi-experimental designs and panel studies create unique challenges. The same is true for our utilization of primary and secondary data analysis. **Secondary data analysis** has a host of strengths that make it a popular approach to social science research.[15] First, some research topics may be impossible or very difficult to analyze using only primary or direct data sources. Say you want to examine public opinion during the Civil Rights Movement in the 1960s. You could try to survey people alive during those years and ask them to reflect on what their opinions were during that time. It is obvious to see the problems with that approach. Respondents may forget important events, or their experiences in the intervening years may distort their responses. A better research strategy may be to examine surveys that were conducted by others during the Civil Rights Movement. Similarly, secondary data allows the researcher to examine cross-national research questions that would be difficult to answer using primary sources. Also, researchers can use secondary source analysis to examine trends over time. Secondary data are often nonreactive. **Reactivity** occurs when subjects act differently because of the study or research setting, and this can bias the results of a research study. Secondary data are often nonreactive because participants, observers and record-keepers are usually unaware of any future research hypotheses. Cost is another important strength of secondary source analysis. Researchers can save a great deal of time and money by using existing data, rather than trying to collect the data themselves.

It is imperative that users of secondary data analysis fully understand how the data they are examining were collected, measured and kept. Researchers must ask if the data are representative of the **population** of interest, and social scientists often face the challenge of incomplete data.[16] Incomplete data may arise from **selective survival**, which means that some documents may survive longer than others. Furthermore, record keepers often do not keep or preserve all records. For

example, if a political scientist is examining the memoirs of presidents, she may notice that negative stories about the presidents are often overlooked and not included in the documents. This is an example of **selective deposit**, which are biases that are related to what type of information is available in documents. If history is written by the victors, researchers must be aware of the possibility of bias in the written record.

Many of the strengths of secondary source analysis are weaknesses of primary source analysis. However, sometimes primary source analysis is better than the alternative. A researcher may have a research question for which data are not readily available. By collecting the data herself, the researcher will be able to ensure that the data will allow her to answer the research question. When using secondary sources of data, researchers must rely on others to measure, code and keep the data.

The Literature Review

For any social science study, it is important to consult the work that previous scholars have published (a **literature review**). This allows us to compare our research question to similar studies in the past, and that helps us assess which methods are more or less appropriate for the specific challenges we face in our own work. By reviewing the literature on school vouchers, we were able to design our study to examine some of the important questions that still exist in the field of school vouchers.

There were two main components of our literature review. First, we examined how previous evaluators of school voucher programs dealt with the issue of **selection bias**. Selection bias can affect voucher evaluations in two general ways. First is the so-called "cream-skimming" hypothesis. According to this argument, those students who are the highest achievers and with the most active and involved parents would be more likely to use a voucher to attend a private school. The second scenario is called "negative selection," and it is the inverse of cream-skimming. Negative selection occurs when those students who are struggling the most in the public system choose to use a voucher. According to

some sources, school voucher programs promote "cream-skimming" by attracting highly educated parents and highly motivated students, while other researchers show evidence of negative selection, in which low-achieving students are more likely to use a voucher. These conflicting results emphasized that we needed to deal with selection bias in our research design.

After deciding to use the matching method to create a sample of public school students that would be similar to our voucher student sample, the research team then had to figure out what factors we would use to perform the match. We used the literature review to find student and parent characteristics that were related to the decision to use a voucher and student attainment. Our analysis of the previous research demonstrated that we needed to account for students' prior achievement scores, race, gender and fluency in English. Our matching process did not stop there, however. We thought that where in Milwaukee a student lived might be important. It could affect students' likelihood of using a voucher, as well as their attainment. We searched the literature and found that neighborhoods could be related to a host of sources of selection bias, like social networks and access to information about schools.

For the second component of our literature review, we looked for gaps in the existing research. Although there have been a number of voucher evaluations, important questions remain. Under the *No Child Left Behind Act* and the subsequent *Every Student Succeeds Act*, public school students are required to take tests in reading and math every year in grades 3 through 8 and once in high school. States, school districts and individual schools, both public and private, often require additional testing of students. These data are widely available to the public, which is one reason they are so common in education program evaluation. Earlier voucher studies focused on student test scores as a measure of the effectiveness of voucher programs. However, we expect schools to do more than just produce excellent standardized test takers. Schools should engender creative thinking, promote citizenship and prepare students to enter the workforce. Given the limitations of standardized tests to measure these outcomes or **dependent variables**, we decided to examine how private schooling, through voucher programs, affects student attainment. While it might seem obvious that graduating from high school and attending college are good outcomes, we needed empirical evidence that provides support for these claims. Therefore, we

examined the literature in economics, sociology, psychology, public health and political science, and we found that increased student attainment is associated with a host of positive consequences, including a longer life expectancy, future wealth, lower crime rates and greater political participation. We were able to use the literature to demonstrate that the dependent variables in our study are important to examine.

Conducting the Study: Collecting the Data and doing the Analysis

Before we could perform our analysis, we needed to select our study participants. There were almost 18,000 students in the MPCP in 2006 when we began our study. We were unable to include all the voucher students, or the whole population, in our study. Since we were following the students in our voucher sample over a five-year period, it would be much too expensive and difficult to include everyone. Therefore, we needed to select a **sample** from the population of voucher students. We wanted to be sure that our sample size would be large enough to account for possible panel attrition if students dropped out of the study. We also wanted to be sure that our sample was representative of the population. If a sample was representative of the population, then we could make inferences about the population, which is generally the goal of this type of political science research. Based on our literature review, we were particularly interested in the effects of the voucher program on high school graduation, college attendance and persistence. As a result, we included all voucher 9th graders in our study. It would be too costly to include all voucher students in 8th grade in the sample, so we **randomly selected** voucher students in that grade. Our final sample size, often denoted by "**N**", includes 1,091 voucher students.

One of our main strategies to deal with selection bias was to match Milwaukee Public School, or MPS, students to our MPCP sample. For each voucher student, we essentially created a list of MPS students in the same grade. Next, using students' home addresses, we identified the census tract in which each student lived. We eliminated from that list all MPS students who did not live in the same

census tract as the MPCP student. Then we looked for MPS students who had similar test scores on their fall 2006 tests. At this point, we often had multiple MPS students who could still be matched to our MPCP sample participant. Therefore, we finally tried to match students based on gender, race and fluency in English.[17] We followed the same matching procedure for each of the voucher students in our sample. By matching on neighborhood, test scores and student demographics, we tried to account for both observed and unobserved selection bias.

How good was our match? Table 9.1 compares our MPCP sample to our matched MPS sample. Looking at the table, one sees that the matched MPS sample looks a lot like our MPCP sample based on these student characteristics. For example, 70 percent of both the MPCP sample and the MPS matched sample are black. The table also includes standardized test scores in reading and math for both groups.[18] The average math test score for the voucher students was below the average math score for MPS students, but the voucher students scored higher than the public school students in reading. These differences were accounted for in our analyses below, but we note that together they provide no clear pattern of performance favoring one sector of our match over the other.

Table 9.1 Comparison of MPCP and MPS Matched Samples

	MPCP 2006	MPS 2006
Black	70%	70%
Hispanic	19%	18%
Asian	3%	4%
White	7%	7%
Female	57%	53%
Math Score 2006	-0.04	0.04
Reading Score 2006	0.15	0.02
N	1,091	1,091

Table adapted from Cowen *et al.* (2013), p. 152

Table 9.2 High School Graduation and Postsecondary Enrollment Rates

Difference

	MPCP in 2006 (%)	MPS in 2006 (%)	Difference
High School Graduation			
On-time graduates			
2006-2007 9th graders	76.0	69.0	7.1[***]
2006-2007 8th graders	73.7	71.6	2.1
Five-year graduates	5.4	9.5	-4.1[***]
Ever graduated	79.0	76.0	3.0
Postsecondary Enrollment			
Two-year university	12.1	14.0	-1.9
Four-year university	25.8	21.5	4.2[**]
Persist in 4-year	21.0	17.9	3.1

[***] $p < 0.01$ [**] $p < 0.05$ [*] $p < 0.10$

Table adapted from Cowen *et al.* (2013), p. 154

After demonstrating that our samples were quite alike when the study started during the 2006–2007 school year, we wanted to examine if graduation and postsecondary enrollment rates differed between voucher and public school students five years later. The results are in Table 9.2.

The table presents the results for a number of different bivariate comparisons. **Bivariate analyses** compare how one **independent variable**, in this case school sector (voucher or public), is related to one dependent variable (e.g., enrolling in four-year university). Looking at Table 9.2, one can examine on-time high school graduation, which means that the student graduated high school when expected (i.e., four years after starting high school). Looking at those who started the study in 9th grade, one sees that 76 percent of voucher students graduated high school four years later, while only 69 percent of MPS students did. However, MPS students were more likely to graduate five years after beginning high school (9.5% vs. 5.4%). Examining both on-time and delayed graduates who were in 9th grade in 2006–2007 and on-time graduates who were in 8th grade in 2006–2007, we find that the total graduation rate for voucher students was 79 percent, and it was 76 percent for MPS students.

Is a three percentage point difference a **statistically significant** difference?

This question involves hypothesis testing. Researchers often examine different levels of statistical significance. In Table 9.2, these are evident at the bottom of the table (e.g., p<0.01, p<0.05, p<0.10). These significance levels, or **p-values**, are associated with different levels of confidence. If you use a p-value of .05, the most common p-value in social science research, this means that there is a 5 percent chance that the relationship you observe between the independent and dependent variable would occur if the two variables were not related at all. If a researcher wants to be very conservative, she can set the p-value very small (e.g., p<0.01); however, this will make it more difficult to find a statistically significant relationship.

When we ask whether the difference we observed is statistically significant, we are asking how likely it is that we could have observed that difference if the true difference were zero. Looking at the difference column for ever graduated, one can see that there is no asterisk, and this means that the difference in total graduation rates is not statistically significant. The lack of a statistically significant relationship between voucher status and graduation rates means that voucher students were neither more nor less likely to graduate from high school than were matched public school students. However, there are statistically significant differences when examining the different components of the total graduation rate. The table indicates that the 7.1 percentage point difference in on-time graduation rate for baseline 9th grade students is statistically significant, as is the difference in delayed graduation for this group. For 9th grade students in 2006–2007, we can conclude that voucher students were more likely than MPS students to graduate on time, while MPS students were more likely to graduate five years after beginning high school. These two opposite findings are one reason why there is not a statistically significant difference when looking at the "ever graduated" row. Further, the results in Table 9.2 indicate that voucher students are 4.2 percentage points more likely to attend a four-year university, and that this difference is statistically significant.

While this result suggests that voucher students may be more likely to attend a four-year university than are public school students, a more sophisticated analysis was needed. To isolate the effect of using a voucher, we used **multivariate regression analysis**. We used it in our study to describe the relationship between participating in the voucher program and our outcomes of

interest, after adjusting for the impact of other factors, like race and gender.[19] The results from our basic analysis are portrayed in Table 9.3. We estimated four regression **models**, one for each of our main dependent variables (graduate from high school, attend a two-year university, attend a four-year university and persist in a four-year college). The goal of the regressions is to isolate the relationship between our main independent variable, which measures if the child was a voucher student for the 2006–2007 school year, and the dependent variable. Students who received a voucher were coded as a 1 for this independent variable, and public school students where coded as a 0. If voucher students were more likely than public school students to graduate high school, for example, the coefficient for the variable will be positive. If public school students graduated at a higher rate than did voucher students, then the coefficient for the voucher variable will be negative.

Table 9.3 Multivariate Analyses: Predicting Attainment Outcomes

	Graduate from High School	Attend 2-Year College	Attend 4-Year College	Persist in 4-Year College
MPCP in 2006	0.13 (0.08)	-0.13* (0.08)	0.20*** (0.07)	0.20** (0.08)
Black	0.07 (0.14)	0.22 (0.15)	0.00 (0.13)	-0.19 (0.15)
Hispanic	0.04 (0.15)	0.01 (0.16)	-0.14 (0.14)	-0.26 (0.17)
Asian	0.99*** (0.27)	0.23 (0.21)	0.17 (0.21)	0.19 (0.25)
Female	0.24*** (0.08)	0.04 (0.08)	0.39*** (0.07)	0.34*** (0.09)
Math 2006	0.19*** (0.06)	0.09* (0.06)	0.17*** (0.05)	0.18*** (0.06)
Reading 2006	0.17*** (0.05)	0.04 (0.05)	0.40*** (0.05)	0.35*** (0.06)
Constant	0.65*** (0.14)	-1.14*** (0.15)	-1.01*** (0.13)	-1.07*** (0.15)
N	1,475	1,830	1,830	1,263

Notes: *** $p<0.01$ ** $p<0.05$ * $p<0.10$. Estimates are probit coefficients. Standard errors are in parentheses. Persistence in 4-year college is restricted to students who were part of the original 9th grade panel. Race variables are indicator variables with "White" as the reference category.

Table adapted from Cowen *et al.* (2013), p. 158.

The results from the multivariate regressions were largely consistent with the bivariate results. We found that voucher students were more likely than matched public school students to attend a four-year college, but they were less likely to attend a two-year college. Further, voucher students were more likely to have persisted in a four-year university for at least one year than were students who attended a public high school. These differences were statistically significant. While the multivariate regression suggests that voucher students might be more likely to graduate from high school because the coefficient is positive, we concluded there was not enough evidence to prove that a meaningful difference exists because the coefficient is not statistically significant.

The biggest advantage of multivariate regression analyses over bivariate analyses is that multivariate regression analyses allow you to examine how multiple independent variables are related to the dependent variable at the same time. With multivariate regression analysis, researchers examine the unique, individual relationship between a particular independent variable and the dependent variable after taking the impact of other independent variables into account. This means that voucher students are more likely to attend a four-year university than public school students are, for example, holding the effects of the race, gender and 2006 test scores constant.

Lessons to Be Learned

As we discussed in the introduction, school vouchers are a controversial topic. As program evaluators, we wanted to provide a transparent and non-ideological analysis of the effects of the voucher program in Milwaukee. Researchers should be aware of the limitations of their study and attempt to account for them in an appropriate way. For example, we realized that selection bias was a serious threat to our analysis. Rather than ignoring it, we tried to adjust for student selection in both the sampling stage through our matching design, as well as in the analysis stage by using multivariate regression analysis. Knowing that your research could be used by policymakers and interest groups to further their political agendas can make this type of research exciting, but also nerve-wracking. It requires program

evaluators to be very careful and transparent in how they describe their research and results.

Given the research question and resources, our evaluation of the Milwaukee voucher program utilized both primary and secondary data sources. One of the greatest strengths of secondary source analysis, which may be of particular importance to political science students, is that it can be quite cost effective. It is often cheaper and easier to use data that have been gathered by someone else, as opposed to trying to collect the data on your own. Furthermore, secondary source analysis can sometimes allow the researcher to examine longitudinal questions or cross-national topics, which could be impossible using other data sources. However, secondary source analysis is not without its weaknesses. Since the researcher is relying on others for data collection, it is often likely that the variables or measurements will not be ideal for the research question.

Official program evaluators have a constituency beyond their academic peers. While many policymakers view evaluations as a way to inform their decision-making, others see evaluations as a way to avoid or postpone difficult decisions. At the same time, legislators often want answers quickly. Finally, policymakers may have goals for the evaluation that are different from what the researcher has. In our evaluation, lawmakers were primarily concerned with the effects of the MPCP on standardized test scores. We believed this was an incomplete approach. In this chapter, we examined student attainment, and in other reports, our research team examined how the MPCP affects parents' views of safety and discipline in their children's schools and how the voucher program has affected taxpayers. It is our view that the effects of a public program should not be reduced to a test score alone.

Performing an evaluation of a public program is often different than other forms of political science research. Legislative mandates and oversight can be quite challenging to researchers. However, we view our participation in this evaluation as a form of public service. According to the Wisconsin Department of Public Instruction, the Milwaukee voucher program currently costs over $200 million per year. Policymakers and taxpayers deserve to know the effects of public policies. Our experiences with the evaluation of

the Milwaukee Parental Choice Program have taught us that political scientists can play an important role in informing citizens and shaping public policy.

Interested to Know More about the Study Discussed in this Chapter?

Cowen, Joshua M., David J. Fleming, John F. Witte, Patrick J. Wolf, and Brian Kisida. 2013. "School Vouchers and Student Attainment: Evidence from a State-Mandated Study of Milwaukee's Parental Choice Program." *Policy Studies Journal* 41(1): 147–168.

Exercises and Discussion Questions:

1. If the researchers had been able to randomly assign students to the Milwaukee voucher program, how may some of the problems with selection bias been avoided? Can you think of any problems associated with selection bias that may have remained? (Hint: Can policymakers force students to attend private schools?)

2. In this chapter, the researchers thought that student demographics, test scores and neighborhood location were important factors that explain the decision to attend a certain school. Think back to your choice of high school. What factors explained your and/or your parents' decision-making process? How could you systematically include those factors in a study like this? Further, what other reasons for attending a particular school may still be unaccounted for in the study described in this chapter?

3. This chapter focuses on the causal effect of attending a private school on student attainment. An important aspect of examining causal effects is understanding the causal mechanisms. Causal mechanisms are "how" or

"why" the independent variable influences the dependent variable. For example, perhaps voucher schools have smaller class sizes, which increases student learning and leads to greater student attainment. Think of some of the possible causal mechanisms that may explain why one type of school could be more effective than another. Consider causal mechanisms that pertain to students, classrooms, teachers, schools, parents or other factors. How could you design a study to examine some of these mechanisms?

4. As part of the Affordable Care Act, many states expanded Medicaid, a means-tested program that extends health insurance to low-income individuals. Outline a potential program evaluation of this Medicaid expansion. What types of outcomes would you want to examine? How could you collect data on those outcomes? How would you design your study (e.g., experimental vs. quasi-experimental)? How might selection bias affect your analysis? Besides selection bias, what are some other challenges of evaluating this program? For examples of Medicaid evaluations, see Decker *et al.* (2013) and Sommers, Baicker, and Epstein (2012).

Recommended Resources:

School Vouchers

SCDP Evaluation Website (www.uaedreform.org/school-choice-demonstration-project/)
News and Research on School Vouchers (www.edweek.org/topics/vouchers/index.html)

Pro-School Voucher Organization (www.edchoice.org/)
Anti-School Voucher Organization (www.nea.org/home/19133.htm)

Reviews of the Research on School Vouchers

Fox, Robert A., and Nina K. Buchanan, eds. 2017. *The Wiley Handbook of School Choice.* Hoboken, NJ: John Wiley & Sons (especially chapters 19–22).

Berends, Mark, Matthew G. Springer, Dale Ballou, and Herbert J. Walberg, eds. 2009. *Handbook of Research on School Choice.* New York, NY: Routledge (especially chapters 14–20).

Examples of Program Evaluation

The World Bank, Development Impact Evaluations (www.worldbank.org/en/research/dime)

U.S. Department of Energy, Office of Energy Efficiency, and Renewable Energy (https://energy.gov/eere/analysis/program-evaluation-examples-and-reports)

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (https://ies.ed.gov/ncee/wwc/)

RAND Corporation (https://www.rand.org/topics/program-evaluation.html)

MDRC (https://www.mdrc.org)

Books and Journals on Program Evaluation

Langbein, Laura, with Claire L. Felbinger. 2006. *Practical Program Evaluation: A Statistical Guide.* London: M.E. Sharpe.

Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation.* 2nd edn. Thousand Oaks, CA: Sage.

Newcomer, Kathryn E., Harry P. Hatry, and Joseph S. Wholey, eds. 2015. *Handbook of Practical Program Evaluation,* 4th edn. San Francisco, CA: John Wiley & Sons.

American Journal of Evaluation, Evaluation and Program Planning, Evaluation Quarterly.

Notes

1 For example, Coleman and Hoffer 1987; Lubienski and Lubienski 2013.

2 EdChoice 2017.

3 In addition to school vouchers, there are a number of other school choice policies that are increasing in popularity, including charter schools, education savings accounts and tax-credit scholarship programs.

4 Witte 2000; Wolf *et al.* 2009; Mills and Wolf 2017.

5 Like some other school voucher programs, the MPCP is targeted at low-income families. In order to receive a voucher for the first time in the MPCP during the years of our study, a family's annual household income had to be 175 percent of the federal poverty level or less.

6 The State of Wisconsin has created three additional school voucher programs. In 2011, a separate voucher program was enacted for students living in the city of Racine, Wisconsin. Second, the Wisconsin Parental Choice Program was authorized in 2013 and is available to parents throughout the state with an income at or below 185 percent of the federal poverty level. The size of this program is capped by state law. Lastly, the state enacted a voucher program targeted at students with special needs in 2015.

7 Langbein and Felbinger 2006: 3.

8 Trenholm *et al.* 2008.

9 Sommers *et al.* 2015.

10 Ennet *et al.* 1994.

11 Shadish, Cook, and Campbell 2002.

12 For examples of the use of field experiments to analyze the effects of school vouchers, see Howell and Peterson 2006 and Wolf *et al.* 2009.

13 For more examples and information on the difference between primary and secondary data, see Cole 1996 and O'Sullivan and Rassel 1999.

14 So, do test prep classes, like those students may take to prepare for the LSAT, SAT, or ACT, actually work? The evidence is mixed, and researchers have been stymied by selection bias. However, a review of the literature suggests that these programs have a "minimal positive effect" on average (Adams 2011).

15 Singleton, Straits, and Straits 1998.

16 Singleton, Straits, and Straits 1998.

17 We used a method called propensity score matching to do this. Propensity score matching involves estimating the probability that a subject is exposed to a particular condition (here, participating in the voucher program) based on a set of observable characteristics about that subject. In our case, these characteristics involved race, gender, prior test scores, and the student's first spoken language.

18 We "standardized" students' test scores to have a mean of 0 and a standard deviation of 1. For our purposes here, what is important is that having a positive average test score meant that the group scored above the Milwaukee Public School District average, while having a negative number indicated the group scored below the Milwaukee average.

19 We used a statistical technique called multivariate probit regression in the article. This technique is used when the dependent variable is binary (e.g., graduated high school (yes/no) or enrolled in 4-year university (yes/no)). One limitation of probit regression is that interpreting the regression coefficients is not straightforward. For our purposes here, a positive probit coefficient means that that group is more likely to have achieved the outcome (e.g., graduate high school or enroll in 4-year university), while a negative coefficient means that the group is less likely to achieve the particular outcome.

References

Adams, Caralee J. "Value of College-Admissions Test-Prep Classes Unclear." *Education Week,* March 2, 2011. www.edweek.org/ew/articles/2011/02/24/22sat-act_ep.h30.html.

Cole, Richard L. 1996. *Introduction to Political Science and Policy Research.* New York: St. Martin's Press.

Coleman, James S., and Thomas Hoffer. 1987. *Public and Private High Schools: The Impact of Communities.* New York: Basic Books.

Decker, Sandra L., Deliana Kostova, Genevieve M. Kenney, and Sharon K. Long. 2013 "Health Status, Risk Factors, and Medical Conditions Among Persons Enrolled in Medicaid vs Uninsured Low-Income Adults Potentially Eligible for Medicaid under the Affordable Care Act." *JAMA* 309(24): 2579–2586.

EdChoice. "Fast Facts on School Choice." https://www.edchoice.org/resource-hub/fast-facts/.

Ennet, Susan T., Nancy S. Tobler, Christopher L. Ringwalt, and Robert L. Flewelling. 1994. "How Effective Is Drug Abuse Resistance Education? A

Meta-Analysis of Project DARE Outcome Evaluations." *American Journal of Public Health* 84(9): 1394–1401.

Howell, William G., and Paul E. Peterson, with David E. Campbell and Patrick J. Wolf. 2006. *The Education Gap: Vouchers and Urban Schools.* Washington, DC: Brookings Institution Press.

Langbein, Laura, with Claire L. Felbinger. 2006. *Practical Program Evaluation: A Statistical Guide.* London: M.E. Sharpe.

Lubienski, Christopher A., and Sarah Theule Lubienski. 2013. *The Public School Advantage: Why Public Schools Outperform Private Schools.* University of Chicago Press.

Mills, Jonathan N., and Patrick J. Wolf. 2017. "The Effects of the Louisiana Scholarship Program on Student Achievement after Three Years." Louisiana Scholarship Program Evaluation Report #7.

School Choice Demonstration Project, University of Arkansas, Fayetteville, AR; Education Research Alliance for New Orleans, Tulane University, New Orleans, LA. June. https://educationresearchalliancenola.org/files/publications/Mills-Wolf-Effects-of-LSP-on-Student-Achievement-After-Three-Years.pdf

O'Sullivan, Elizabethann, and Gary R. Rassel. 1999. *Research Methods for Public Administrators.* 3rd edn. New York: Longman.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Singleton, Jr., Royce A., Bruce C. Straits, and Margaret Miller Straits. 1998. *Approaches to Social Research.* 2nd edn. New York: Oxford University Press.

Sommers, Benjamin D., Katherine Baicker, and Arnold Epstein. 2012 (September 13). "Mortality and Access to Care among Adults after State Medicaid Expansions." *New England Journal of Medicine* 367: 1025–1034.

Sommers, Benjamin D., Munira Z. Gunja, Kenneth Finegold, and Thomas Musco. 2015. "Changes in Self-Reported Insurance Coverage, Access to Care, and Health under the Affordable Care Act." *JAMA* 314(4): 366–374.

Trenholm, Christopher, Barbara Devaney, Kenneth Fortson, Melissa Clark, Lisa Quay, and Justin Wheeler. 2008. "Impacts of Abstinence Education on Teen Sexual Activity, Risk of Pregnancy, and Risk of Sexually Transmitted

Diseases." *Journal of Policy Analysis and Management* 27(2): 255–276.

Witte, John F. 2000. *The Market Approach to Education: An Analysis of America's First Voucher Program.* Princeton, NJ: Princeton University Press.

Wolf, Patrick J., Babette Gutmann, Michael Puma, Brian Kisida, Lou Rizzo, and Nada Eissa. 2009. *Evaluation of the D.C. Opportunity Scholarship Program: Impacts after Three Years.* Washington, DC: U.S. Department of Education.

Content Analysis

Congressional Communication Through Broadcast and New Media

C. Danielle Vinson

<div style="border:1px solid black">

CONTENTS

</div>

Getting Curious

After ten years of research and writing, my book manuscript on how members of

Congress use the media to shape policy was finished, or so I thought. Then the email from the publisher arrived, telling me that the editor liked the book proposal and sample chapters but he wanted me to include a chapter on television and social media. Ugh! That was in 2012, and I was the person who still had no Facebook account and was quite dubious that anyone could communicate anything of significance in 140 characters or less on Twitter. And now I was being told I had to conduct research on Congress and social media if I wanted my book to be published. Sometimes we get curious and do research because *we* want to know something; other times we do it because *our editor* wants to know something. If it meant my book would be published, I was willing to be curious about Congress and social media, though I reserved the right not to enjoy it.

The editor was not wrong. In my book, I was looking at how members of Congress have tried to use the media to influence policymaking and how that has changed over the last 40 years. Social media is arguably one of the most important communication developments during that time because it gives members of Congress a consistent way to alert both the media and the public to their concerns and positions on issues. And unlike mainstream media – newspapers and network and cable television news – social media have no gatekeepers that members of Congress have to satisfy to get their ideas or stories published. Through social media, members can talk about anything they want to discuss. And in 2012, political scientists knew almost nothing about how members of Congress were actually using social media because there was little research on it.

The Research Strategy: Content Analysis

Recognizing the importance of the issue and the necessity of satisfying my editor, I now needed to put together a **research design** or strategy for studying my question. For me, the method of research was quite obvious – **content analysis**. To understand how members of Congress use social media, I needed to *systematically* examine and analyze members' social media posts.

Content analysis is a way to use categories to systematically analyze written

or broadcast records. The process allows us to convert verbal or visual non-quantitative (non-numerical) content into data that can be measured quantitatively. We first choose what content we want to analyze. Almost any written, oral or visual material can be subjected to content analysis, including media coverage, speeches, press releases, political ads, cartoons and photos – but the content should be appropriate for investigating the researcher's question. In most cases, we do not have enough time or sufficient manpower to analyze all the material that might be relevant to our research question, so we must choose a subset of the material – what we call our **sample**. For example, I could not examine all the tweets, Facebook posts and videos on YouTube from every member of Congress, so I had to pick a small number over a specific time period to analyze.

Once we have selected what we want to analyze, we create **content analysis categories** that measure the aspects of the content we want to investigate. For example, suppose we want to understand how partisan congressional members are on Twitter. We would select as our content a sample of tweets from a sample of members. To construct our categories, we would consider what aspects of a tweet might make it partisan. Thus, our categories might include several measures. First, we might identify the targets of the tweet and its reaction to those targets: is it supporting or opposing the president or members of either political party? Second, we could look at the purpose of the tweet – was it about taking a position on an issue, which might highlight partisan disagreements, or constituent outreach or the member's personal life, which would not be partisan. We could also look at the issue of the tweet because we know that certain issues tend to belong to one party more than another – something we call "party ownership" in political science. Together, these aspects of the coverage might help us determine how partisan members of Congress are in their communication on Twitter.

Table 10.1 Partial Content Code for Congressional Members' Tweets

A. Issue in the tweet	B. Purpose of the tweet
1. Budget/taxes	1. Position-taking
2. Economy	2. Advertising
3. Business	3. District outreach

4. Foreign policy	4. Credit-claiming
5. Defense	5. Providing other information
6. Healthcare	6. Response to other Tweet
7 Social policy	7 Personal
8. Environment	**C. Reaction to President**
9. Energy policy	0.No
10. Crime	1. Oppose/criticize
11. Agriculture	2. Support
12. Education	**D. Reaction to Democrats**
13. Civil rights	0.No
14. Government affairs	1. Oppose/criticize
15 Appointments	2. Support
16. Party politics	**E. Reaction to Republicans**
17 Immigration	0.No
18. Transportation	1. Oppose/criticize
19. Other	2. Support

Choosing categories and defining them requires careful thought and planning and a fair amount of trial and error. Categories need to be both exhaustive and mutually exclusive. Exhaustive means that the categories must cover all possibilities, and mutually exclusive means all the content must be able to be assigned to a single category and cannot fit more than one. Suppose we want to look at the partisan tone of a congressional member's tweets. Can tone only be partisan or not partisan? What would we do with a tweet that criticizes *both* parties? We would need more categories to ensure that they are exhaustive and cover all possibilities for the tone of the tweet. Continuing this same example, we might also want to know the targets of the tweets to help us determine which targets are more likely to be supported or opposed by the congressional member. We might create categories of possible targets that include people in the member's own party, people in the opposing party and the president. The potential problem with these categories is that the president falls into two categories. He is president, but also is either in the congressional member's party or the opposition party. The categories are not mutually exclusive. We could

solve the problem by clarifying that the party categories mean anyone other than the president who is in the congressional member's party or opposing party.

Once we have created our categories, we need to determine our **unit of analysis** – that is, the level of the content we will analyze. For example, if we are studying news coverage of Congress, we can analyze an article as a whole, or we can look at each paragraph or sentence. The unit of analysis will depend in part on what we want to know and the categories we have created. If we want to know what makes coverage of Congress negative, it makes sense to use the story as the unit of analysis. But if we want to know if one member is covered more negatively than another, it might be better to code each sentence of the story rather than the story as a whole.

Having chosen the content we wish to study, created and defined the categories for doing so, and selected our unit of analysis, we need to compile the categories in a **content code** (see [Table 10.1](#)). We can use the content code to analyze each unit of analysis, recording the answers for each in a spreadsheet or appropriate statistical program. Once we are done, we will have a full record of how often each category occurred in our content sample, and we can begin to perform a number of statistical operations described in more detail later in the chapter.

Issues with this Research Strategy

Content analysis offers several advantages to researchers. First, it can be applied to a wide variety of content and messages. We can analyze not only the written and spoken word from sources including news media, government documents, speeches and television shows, but we can also examine visual content in photos, cartoons, television, film and art. For example, some scholars of Congress have examined the photos and mementos that congressional members display on the walls in their Washington offices to see what messages they are communicating to visitors to their offices. The ability to analyze so many different types of content is particularly valuable when studying new media such as a congressional member's website or social media because they combine written

and spoken words, visual images and non-verbal sounds. Furthermore, the applicability of content analysis to such a wide range of content makes it useful for studying nearly any subject, not just American government. In International Relations, it has been used to study foreign policy leadership.[1] One well-known example from Comparative Politics is the Comparative Manifesto Project – a content analysis of European political parties' election programs since 1945 that provides data to study parties within specific countries or across countries.[2]

Second, many of these sources of content are increasingly easy and inexpensive to access and search. Many newspapers from around the world are readily available on the internet and are archived in electronic databases. Likewise, abstracts (summaries) or transcripts of television news shows, political talk shows and radio shows are also available online. The internet has expanded our access to transcripts of government hearings, legislative debates and presidential speeches. Elected officials often maintain their own websites where they post their speeches, clips of radio and television interviews and press releases, and of course, they have social media accounts.

A third major advantage of content analysis is its suitability to both **quantitative** and **qualitative analysis**. The primary purpose of content analysis is to transform written or visual non-quantitative data into quantifiable categories that make quantitative, or statistical, analysis possible. As we count the occurrence of each category in our units of analysis, we can determine the **frequency**, or how often something occurs. For example, if we look at campaign ads, we can determine whether the tone of each is negative or positive. When we are done coding our entire sample of ads, we can see how many were negative. Content analysis also allows us to see how often certain categories are **correlated** with each other or other variables – that is, how often they occur together – and how they co-vary – whether they increase or decrease as other categories or variables increase or decrease. We might, for example, look to see if the number of negative ads in a congressional race is higher in those races that are close or if they increase as Election Day gets closer. Depending on the nature of our categories, we may be able to perform sophisticated statistical analysis on our data, as we will see later in this chapter.

But content analysis is not limited to **quantitative analysis**. It can also enable us to do more in-depth qualitative examination. For example, while we can look

at political ads to see how often candidates go negative, we can look at qualitative categories to analyze what makes an ad negative. We might look at the specific language or images candidates use. The ability to extract qualitative data and examples from content analysis often helps to illustrate the patterns we find in the quantitative analysis and makes our presentation of it more interesting.

Despite the benefits of content analysis, it is not without its costs and challenges. One important concern is the **validity** and **reliability** of our categories. Validity is whether the categories measure what we claim they do. For example, when we code news content as negative or positive, are we measuring how the story was covered or the coder's opinion of the issue being covered? Our content code needs to be clear about how we have defined negative and positive to insure validity. Reliability is whether someone else following our definitions and procedure would arrive at the same coding decisions. In content analysis, these two issues are often connected. For objective categories, such as the date a news article appeared, the measure is clearly valid and reliable. But on more subjective matters such as the tone of a news article, there may be more room for the coder's own judgment to play a role. Researchers must be careful that their own selective perception – the tendency to filter and interpret what we read or hear through our own ideological preferences – does not bias the analysis. Suppose we are analyzing the tone of a news article that reports that "Republicans plan to try to stop passage of the President's bill." Is that statement favorable or unfavorable to Republicans, or is it neutral reporting? If my analysis depends on whether I like the President's bill, then we have a problem with validity and reliability because I am now reading my own views into what the reporter actually said (a problem for the validity of the measure), and I cannot be certain that another person would arrive at the same interpretation (a problem for reliability).

Content analysis must provide clear definitions of categories that can be applied consistently and with the same results regardless of who does the coding. To guard against subjective measures, we typically conduct **inter-coder reliability** tests. We enlist a second person to code a small sample of our data and measure the level of agreement or discrepancies between the two coders. On objective categories, we would expect complete agreement. On subjective

categories, we look for agreement on more than 80 percent of the cases. If inter-coder reliability falls below 80 percent, it may be an indication that the categories and definitions are problematic and need further clarification.[3]

One additional challenge in content analysis is the practical problem that this method can be labor intensive and time consuming. It takes time to search databases to find the appropriate content to analyze, and the actual analysis can take even longer. In my own research on congressional members' use of television to try to influence policy, I had to search 19 years of transcripts from the Sunday morning talk show *Meet the Press* to identify every interview with a member of Congress on the show. Then I had to read the transcripts for the 915 interviews with congressional members and code five aspects of each member's comments. Many scholars who do large-scale content analysis rely on graduate students to help with the coding. Increasingly, scholars are turning to computer-assisted content analysis.

Great strides have been made in software programs that aid in content analysis. Researchers can download the material they want to code and create a lexicon or dictionary that the computer will then apply to the downloaded content. Essentially, the computer searches for the words, phrases or concepts of interest to the researcher and counts how often they occur. These innovations are useful and can exponentially increase the amount of content scholars can analyze, but they too have limits, particularly for studies in which the context of a word or phrase is important. For example, computer-assisted content analysis would be quite helpful for learning how media references to the pro-life and pro-choice movements have evolved over time by counting the occurrence of labels such as "pro-life," "anti-abortion," "pro-choice" or "pro-abortion" from media coverage over a long period of time. But this would not necessarily tell us whether the coverage of the movement is favorable or unfavorable or who actually used these labels to refer to the movement. That might require more labor-intensive human coding.

Even if we are able to overcome the challenges posed by content analysis, we must recognize its limits. Analyzing content alone may not help us explain the causes of content or determine its effects. For example, we might study news coverage of a member of Congress and find that it is mostly unfavorable and that the member is rarely quoted in the coverage. We might conclude that the press

has treated the member unfairly. But there is an alternative explanation – the member might not return reporters' phone calls, leaving the reporter with only one side of the story. We have no way of knowing which explanation is accurate from the content analysis alone. If we want to explain the content or learn the effects it has on readers or viewers, we must combine content analysis with other methods such as elite interviews or surveys of public opinion. This use of multiple methods to study a research question is called **methodological pluralism**.

For my own research question, content analysis was the only method that could help me understand how members of Congress use social media. But before I could make decisions about what media to analyze and what categories to include in my analysis, I needed to conduct a literature review of the existing scholarly research that was relevant to my topic.

The Literature Review

The existing research on social media and Congress was sparse and limited in its scope. After all, at that point congressional members had really only been using Twitter and Facebook in significant numbers for three or four years. There were no books on the specific subject, and the research that was available looked primarily at Twitter, not other forms of social media.[4] What I discovered was not so much gaps in the literature but a big gaping hole in a rapidly developing area of political communication.

The research that did exist focused first on which members of Congress were using Twitter. There was a consensus that Republicans posted to social media significantly more frequently than Democrats did.[5] Some scholars suggested that this was primarily because Republicans were the minority party in Congress at the time, rather than anything inherent to Republicans: being in the minority might have made them feel the need to employ more innovative ways to communicate because traditional media tend to be more interested in the majority party, which has more power to influence policy.[6] The research was at odds over whether there were differences in social media use between the Senate

and House. Some found no differences,[7] while others studying a slightly different time period found that senators posted to social media more frequently than representatives.[8] The discrepancy suggested more research was needed.

The second focus of the literature was on why members used social media and what they were communicating through it. Most of the research argued that members had mostly adapted their old messages to the new technology. Three familiar kinds of communications – advertising their names to constituents, claiming credit for their accomplishments and taking positions on issues – were consistent among members' tweets and posts.[9] In addition, members provided information about the congressional district or state or issues. The existing research also noted that members rarely focused on personal information in their social media messages.

Surprisingly absent from the discussion was any attention to partisan messages in members' social media communications. Most research on Congress found partisanship to be an important component of the current congressional environment.[10] Indeed, a recent book on congressional communication entitled *When Politicians Attack!* discussed the fact that congressional communication in traditional media typically aimed at attacking the opposing party.[11] This was an area my own research would need to address.

Conducting the Study: What Coverage and what Categories?

The first issue I had to resolve was what social media I would use for my analysis. I wanted to include a variety of social media from a random sample of congressional members, but I needed to keep the sample manageable. The congressional tweets for all members of Congress were aggregated in a live feed on the website *Tweet Congress*, making it easy to capture all the tweets coming out of Congress at any given period of time.[12] Whichever members tweeted during the time periods I chose, would be the ones I analyzed. Unfortunately, there was no similar collection of Facebook posts or YouTube videos; I would have to go to each congressional member's Facebook page or YouTube channel to

analyze their posts and videos, a time-consuming task even if I limited my sample to a fraction of the 535 members of Congress. For that reason, I decided to focus the content analysis on a random sample of tweets from congressional members and do more qualitative case studies of the Facebook pages and YouTube channels of just a few members of Congress. Therefore, I will focus my discussion in this section primarily on the content analysis of Twitter.

The next issue was to figure out the time period for following tweets. *Tweet Congress* provided a live feed of tweets but no archive, making retroactive retrieval of tweets too difficult and time consuming. I would have to pick time periods in the future rather than in the past. I wanted to insure that the period I studied would be typical of Congress, but with no crystal ball to see into the future, I had to find a way to minimize the likelihood that I would pick a time period that turned out to be atypical. The solution was to pick nonconsecutive days over a couple of months. My research assistant and I settled on one daytime hour on nine nonconsecutive days in February and March of 2013. The times of day and days of the week were varied to insure we would capture the full range of congressional activities in Washington and back home in members' states or districts. Sometimes, even when retroactive retrieval is possible, using nonconsecutive days or weeks can be a good tactic for minimizing the effects of a single event or issue and providing a more representative and manageable sample of content.

Although technology and electronic databases have made retroactive retrieval of content easier and less expensive, researchers still have to make tough decisions about what content and time periods they will study. If we fail to make choices to limit the sample, the content analysis can quickly become overwhelming and the data unwieldy.

My plan was to analyze all the tweets that appeared on *Tweet Congress* during the time periods I had chosen. Because tweets are short with only 140 characters, it made sense that the entire tweet should be the unit of analysis. For longer content – Facebook posts, for example – I would have needed to determine whether smaller units of analysis such as paragraphs or sentences within the post would have been more appropriate than the post as a whole. When my research assistant and I finished gathering tweets, we had 324 to analyze.

The next step was to create the content code. I had to devise categories to

measure the aspects of the tweets I was most interested in – who was tweeting, the substance of the tweets, and the purpose of the tweets. For each aspect I wanted to study, I had to create a list of categories that would encompass all possible options of that aspect of the tweet (see [Table 10.1](#)). Content codes from existing scholarly research offered a starting point for creating categories for some parts of my analysis. I could modify them to fit my particular interests rather than starting over from scratch. That would also make it easier to connect my work to the existing research.

Who was tweeting was easy; I could just record the name of the member of Congress who had posted the tweet. That allowed me to go back later and add useful information about each member: their party and whether they were in party or committee leadership positions; the house of Congress they were in; their ideology (very liberal to very conservative); their age, sex and race or ethnic identity; and the percent of the vote they had won in their last election as a measure of their electoral vulnerability.

To capture the substance of the tweets, I needed to consider several aspects of the content – the issues and the targets of the tweets and whether the targeted messages were partisan or not. For the issues, I could borrow from my own research and use the categories I had created for the content analysis of news coverage of Congress in the earlier chapters of the book. In that analysis, I had created 19 categories of broad issue areas through trial and error. For example, I began with budget and taxes as two separate categories only to discover that these were nearly inseparable in congressional conversations. Because they caused considerable confusion that led to low inter-coder reliability scores, I combined them into one category, eliminating the problem. These issue categories worked well for the tweets except that I found myself putting one third of the tweets in the "other" category that was meant for anything that did not fit in the 18 specific issue areas. In my analysis of news coverage, only 2 percent of the 6,000 stories had been categorized as "other." That suggested that I might need to re-examine that category and see if there were any additional issues that should be broken into their own separate categories. My research assistant and I reread the tweets we had coded as "other" and could find no consistent issues areas among them: they were quite random – everything from members commenting on the weather to their favorite sports to what they had for lunch.

We decided to leave the categories as they were. If I were just beginning this project today, the comprehensive issue categories used in the Policy Agendas Project, now available online, would be a useful starting point, though I imagine it too would have difficulty sorting the random "other" topics in many of the tweets.[13]

In addition to the issues, I wanted to know about any partisan messages in the tweets, given the overall impact of partisan polarization on Congress in recent years. To do this, I looked at the targets of the tweets. I created categories for three possible targets: the president, others in the member's own party, and others in the opposing party. For each possible target, there were three options: 1) the tweet did not explicitly mention that target; 2) the tweet supported or reacted favorably to the target; 3) the tweet opposed or reacted negatively to the target. This would allow me to see not only who members talked about on Twitter but whether they did so in partisan ways (opposing those in the opposite party or supporting those in their own party).

Finally, I wanted to consider the reasons members of Congress tweet. Here, the existing research was very helpful. Most of the earlier studies had been curious about the purpose of tweets and posts and had already developed content categories for this. I borrowed from the existing content codes to create seven categories that identified the purpose of the tweets. *Position-taking* involved members taking a stand on an issue. *Advertising* tweets announced members' appearances in the media or in the state or district. *District outreach* announced district events or accomplishments of those in the district and invited constituent feedback or participation. *Credit-claiming* touted members' accomplishments. Some tweets *provided information* without the member taking a position. *Response* tweets retweeted or directly commented on someone else's tweet. And *personal* tweets included anything that was not related to congressional members' work.

Once I had coded all the tweets and posts, I could begin to analyze the data. First, I wanted simply to describe what members were communicating via Twitter. I did this by conducting frequency analysis of each category – that is, how often it occurred in the tweets or how many of the tweets in my sample included a specific category. For example, I conducted a frequency analysis on the issues in the tweets. I found that in 33 percent of the tweets the issues fell in

the "other" category, 25 percent in the budget and tax category, and 15 percent in crime (due mostly to the debate and passage of one crime bill during the time period I studied). None of the other issues appeared in more than 5 percent of the tweets. This confirmed my expectation that members of Congress would feel free to explore a wider range of issues on social media than they are able to address in traditional media where they must get past the gatekeeping journalists who tend to limit discussion to a handful of issues of national importance.

Because I was particularly interested in determining if the partisan polarization that has affected most aspects of Congress in recent years had carried over into members' social media messages, I also looked at the frequencies of the targets of the tweets. From my own research for the book, I had learned that members' communication in traditional media had become extremely partisan in recent years and often targeted politicians and parties very directly. Therefore, I was surprised to discover that only 9 percent of the tweets explicitly mentioned the president or members of either party. But of those that did include a partisan target, 86 percent expressed opposition to that person or party. Of course, opposition to a politician does not necessarily indicate partisanship. It could have been members of the president's own party criticizing him. While the frequencies allowed me to know members had opposed a politician or party, I needed to know if the member who tweeted was targeting politicians in the opposite party or those in his or her own party. For that I needed to conduct more complex analyses.

Crosstabulations, sometimes referred to as crosstabs, would allow me to see if categories that described targets of the tweets related to the variables I had added to describe the party identification of the congressional members as I expected they would. Crosstabs look at two or more variables to see how often they occur at the same time. For example, I looked at whether a tweet that opposed the president was written by members of his own party or by those of the opposing party. The results were as I expected: opposition to the president or the parties in the tweets came exclusively from members of the opposing party, and support came from those in the same party.

At this point in my analysis, I had enough research on social media to satisfy my book editor, but a funny thing had happened. I discovered that *I* wanted to know more about partisanship in congressional members' tweets. Unfortunately,

my sample had too few partisan tweets to do much analysis beyond crosstabs. The wonderful thing about research, however, is that you do not have to learn everything in one project. I decided I would pursue my study of Congress and social media beyond what I needed to do for the book. With help from students in my American Government classes, I created a new sample of social media that included more than 4,000 tweets and Facebook posts from 154 members of Congress, and I used the same content code to analyze the new data. This has allowed me to understand how the likelihood of members supporting or opposing a particular party in their messages is connected to several characteristics of the congressional members.

To do this analysis, I ran a series of **logit regressions** – a statistical method that allows us to see the impact of several independent variables on a dependent variable that is dichotomous or has only two possible values. In my first regression, the dependent variable, the aspect of the social media message I wanted to explain, was whether the tweet or post had opposed or supported a Democrat. The case was coded 0 if it opposed and 1 if it supported a Democrat. The second regression used whether the tweet or post had opposed or supported a Republican as the dependent variable. The independent variables, what I thought would explain the dependent variable, in both regressions included the members' party identification, whether they were in the Senate or House, their sex and race, and whether they held a leadership position in the party or in a committee. To measure each of these characteristics, I used **dummy variables** – a variable that equals 1 if it is in a particular category and 0 if it is not. So, if a member was a Republican, he or she was coded 1, and if not, I coded 0. If the member was in the Senate, I coded 1, and if not, I coded 0 and so on for each of the independent variables. The results appear in .

The regression analysis revealed that members of a party are more likely to support their own party and oppose the other party, and that senators were more likely than House members to be supportive of either party. I also learned that women were more likely than men to support the Democrats in their tweets, but there was no significant difference between men and women when it came to support or opposition to the Republicans. We know this because the coefficients for these variables were all **statistically significant**. Their **p-values** were all smaller than .05, meaning the probability of these patterns in the independent

variables and the dependent variables occurring randomly if there were no relationship between the variables was less than 5 percent. We can also see in the parentheses in each cell in Table 10.2 the **first differences** or the likely increase or decrease in the dependent variable if we raise each independent variable from its minimum to maximum value while holding the other variables constant. Thus, going from a Democrat to a Republican decreased the likelihood that the tweet or post would support rather than oppose a Democrat 98.6 percent, all else being equal. The likelihood of supporting rather than opposing Democrats increased 770 percent if the member was a Senator rather than a House member, and senators were more than 1,700 percent more likely to support Republicans than House members. Clearly, there was little praise for either party from House members.

There are several numbers below Table 10.2 that are commonly reported in statistical analysis. **N** is the number of cases used in the analysis – in my study that was the 286 tweets and posts that explicitly targeted Democrats and the 318 that targeted Republicans. Below that is the **R-squared** (R^2) **statistic**. It tells us the goodness of fit of our model or how well the independent variables account for or explain the variation in the dependent variable. In this case, the R^2 suggests that about 52 percent of the variance in the reaction to the Democrats is explained by the independent variables and 63 percent for the reactions to the Republicans.

Table 10.2 Congressional Members' Reactions to the Parties on Social Media

Independent Variable[a]	Reaction to Democrats[b]	Reaction to Republicans
Republican	-4.240[*]	6.311[**]
	(-.986)	(549.800)
Senate	2.164[***]	2.917[***]
	(7.702)	(17.477)
Women	1.820[*]	.289
	(5.172)	(.335)
African American	1.127	-4.294
	(2.085)	(-.986)
Committee Leader	.020	.905

	(.020)	(1.471)
Party Leader	.558	-1.259*
	(.746)	(-.716)
Constant	2.944*	-2.424*
	N=286	N=318
	R^2=.507	R^2=.631

* Significant at p<.05; ** Significant at p<.01; *** Significant at p<.001;

a The dependent variable for each column was whether the tweet or post supported or opposed the political party (0 if opposed, 1 if supported).

b The first number in each cell is the logistic regression coefficient. The number in parentheses is the first difference. First differences indicate the increased or decreased likelihood of the dependent variable if we raise the independent variable from its minimum to its maximum while holding the other variables at their mean.

Overall, my study for my book has moved us toward a more complete picture of social media use by members of Congress and allowed us to compare their social media messages to their communication through traditional news outlets. Members do use new media for many of the same purposes for which they use traditional media. However, the lack of gatekeepers on social media has allowed members of Congress to address a much broader range of issues than they can in traditional media. And somewhat surprisingly, we find that most social media messages are not blatantly partisan, in contrast to much of members' comments in traditional media. My research has also raised questions for future scholars to explore.

Lessons to Be Learned

Apart from the substantive findings, there are several things we can take away from this study. First, in the realm of content analysis, we must accept the fact that we cannot control everything in the research (in my case, events or issues dominating the time periods I studied or types of social media I could not easily aggregate and collect). We need to acknowledge our

deficiencies and try to minimize their effects.

Second, I learned that quantitative and qualitative analysis can be complementary. In my case, much of my analysis was quantitative, but systematic qualitative examination of the tweets often compensated for the limitations of my quantitative measure of content and made the research much more interesting. For example, to supplement my assessment of the partisanship in the tweets, in addition to looking at the political targets, I did a qualitative examination of the way the issues were presented in the tweets to determine if the framing favored one party or the other. This highlighted another way that partisanship could creep into social media messages.

Finally, I found that spending time exploring a topic I was not initially interested in raised my curiosity. It opened my eyes to the ways politicians adapt to new technologies and made me curious about what politicians choose to communicate when they have no one telling them what is or is not newsworthy or what is or is not interesting. It turns out that social media can be a window into the minds and personalities of politicians. And much to my surprise, I actually enjoyed the research.

Now, look around you. There is a world of content that has an impact on politics and society. What research questions can you think of that would be appropriate to study with content analysis?

Interested to Know More about the Study Discussed in this Chapter?

Consult the research publication:

Vinson, C. Danielle. 2017. "Chapter 4: New Paths to Influence: Broadcast and New Media." In *Congress and the Media: Beyond Institutional Humor*. New York: Oxford University Press.

Exercises and Discussion Questions:

1. What are the weaknesses of content analysis, and how might you overcome these problems?
2. Try to devise a way of determining whether content is negative or positive that is not dependent on your own ideology or preferences.
3. What challenges might new media (websites, Twitter, Instagram, or other social media) pose for doing content analysis that we don't encounter with traditional media?

Recommended Resources:

On content analysis:

The Content Analysis Guidebook (http://academic.csuohio.edu/kneuendorf_ka/content/): An online companion to *The Content Analysis Guidebook* by Kimberly Neuendorf. It includes content analysis resources, bibliographies of studies that use content analysis, and links to message archives that could be used for content analysis.

Sources of content that could be analyzed:

[PresidentialRhetoric.com](http://presidentialrhetoric.com/) (http://presidentialrhetoric.com/): Transcripts of recent and historic presidential speeches.

The American Presidency Project (www.presidency.ucsb.edu/sou.php): Transcripts of State of the Union addresses, recordings of presidential radio addresses and a wide range of presidential documents including signing statements and executive orders.

Vanderbilt Television News Archive (http://tvnews.vanderbilt.edu/): Abstracts of all network television news from 1968 to present.

Tweet Congress (http://tweetcongress.org): Provides information about which members of Congress use Twitter and links to their Twitter pages. It includes livestream channels of tweets broken down by Senate, House, Congress and both parties.

Manifesto Project (http://manifestoproject.wzb.eu): Provides content analysis of European party election programs since 1945. The website provides codebooks and data for the project.

Notes

1 Walker, Malici, and Schafer 2011.

2 "Manifesto Project Database."

3 Lombard, Snyder-Duch, and Campanella Bracken 2010; Neuendorf 2002.

4 See, for example, Glassman, Straus, and Shogan 2010. One notable exception to the exclusive focus on Twitter was Lawless 2011.

5 Glassman *et al.* 2010; Lawless 2011; Shogan 2010.

6 Lassen and Brown 2011.

7 Glassman *et al.* 2010.

8 Lassen and Brown 2011.

9 Lawless 2011; Shogan 2010; Golbeck, Grimes, and Rogers 2010.

10 See, for example, Sinclair 2006.

11 Groeling 2010.

12 "Tweet Congress."

13 "Policy Agendas Project."

References

Glassman, Matthew, Jacob Straus, and Colleen Shogan. 2010. "Social Networking and Constituent Communications: Member Use of Twitter during a Two-Month Period in the 111th Congress." *Congressional Research Service.* https://www.fas.org/sgp/crs/misc/R41066.pdf.

Golbeck, Jennifer, Justin Grimes, and Anthony Rogers. 2010. "Twitter Use by the U.S. Congress." *Journal of the American Society for Information Science and Technology* 61(8): 1612–1621.

Groeling, Tim. 2010. *When Politicians Attack! Party Cohesion in the Media.* New York: Cambridge University Press.

Lassen, David S., and Adam R. Brown. 2011. "Twitter: The Electoral Connection." *Social Science Computer Review* 29(4): 419–436.

Lawless, Jennifer L. 2011. "Twitter and Facebook: New Ways to Send the Same Old Message?" In *iPolitics*, ed. Richard L. Fox, and Jennifer Ramos. New York: Cambridge University Press, 206–232.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2010.

"Practical Resources for Assessing and Reporting Inter-coder Reliability in Content Analysis Research Projects." http://astro.temple.edu/~lombard/reliability/#top.

"Manifesto Project Database." http://manifestoproject.wzb.eu/.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook.* Thousand Oaks, CA: Sage.

"Policy Agendas Project." www.policyagendas.org/.

Shogan, Colleen. 2010. "Blackberries, Tweets, and YouTube: Technology and the Future of Communicating with Congress." *Congressional Research Service*: 231–233.

Sinclair, Barbara. 2006. *Party Wars: Polarization and the Politics of National Policy Making.* Norman, OK: University of Oklahoma Press.

"Tweet Congress." www.tweetcongress.org/home.

Walker, Stephen, Akan Malici, and Mark Schafer, eds. 2011. *Rethinking Foreign Policy Analysis: States, Leaders, and the Microfoundations of Behavioral International Relations.* New York: Routledge.

Field Experiments: Wired to Mobilize

The Effect of Social Networking Messages on Voter Turnout

Holly Teresi and Melissa R. Michelson

CONTENTS

Getting Curious: Can Facebook Posts Increase Voter Turnout?

If you're an active Facebook user, you've seen plenty of opportunities to click to show your agreement with a movement, to show solidarity with victims of a tragedy, or, on Election Day, to proudly tell your friends "I voted." You might have wondered whether your clicks are influencing other people – making them more likely to click the same buttons, or, in the latter example, making them more likely to vote. Can messages spread on social media influence voter turnout? That question is the one explored in this chapter.

The legitimacy of a democracy depends on active participation by the governed – the voters. Yet, despite widespread acceptance of the idea that voting is a civic duty and normatively desirable, most eligible Americans are not regular voters. Low voter turnout, particularly when active voters are not representative of the broader population, calls into question the degree to which elected officials and public policy represent public needs and preferences. Thus, voter turnout is of interest for both the maintenance of American democracy and also to ensure political equality. These concerns underlie the longstanding interest in participation and turnout by political scientists, including the recent surge of interest in conducting randomized field experiments to test ways of increasing participation among nonvoters.

The growing mobilization field experiments literature generally concludes that face-to-face and live telephone interpersonal communication increase participation, while electronic, mass media and mailed communications, including messages delivered via robocalls, mailers, leaflets, email, text message, television or radio, often referred to as indirect methods, tend to have weak to negligible effects.[1] At the same time, there is reason to believe that online social networking sites (SNS) like Facebook could effectively be used to increase voter turnout, particularly among younger voters.

Inspired by Robert Putnam's groundbreaking 2000 book *Bowling Alone*, a number of social scientists have since sought to explore how to build social

capital and encourage civic engagement and political participation. Much of the social capital literature builds on Putnam's argument that face-to-face interaction is what builds community and trust, which in turn stimulates political participation. Yet, an increasing amount of interpersonal interaction is conducted via social networking sites (SNS) such as Facebook, Twitter and LinkedIn. Research with student samples and broader survey data finds a statistically significant relationship between intensity of SNS use and political participation, both online and offline, suggesting that the increased use of online social networking sites should not be interpreted as a danger to social capital but rather as an alternative means of generating it. This inspired our decision to try to use Facebook posts to increase voter turnout among undergraduate students, in an experiment conducted at a large public university in 2010.

The major **hypothesis**, or expectation, about the nature of the relationship between our **independent** (causal) and **dependent** (effect) **variables** investigated in this study is that social network sites provide an environment and structure for relatively indirect get-out-the-vote messages (our independent variable) to generate significant increases in turnout (our dependent variable). Specifically, the study tests the hypothesis that randomly assigned exposure to political SNS messages that encourage voting in the November 2010 election, **operationalized** as, or measured by, Facebook status messages, will increase the likelihood of an individual voting in that election. Previous scholarship finds that SNS members conceptualize people they accept in their online network as members of their in-group, even if they would not be conceptualized that way offline.[2] Repeated exposure to status updates and posts from these online friends creates a sense of closeness and familiarity that otherwise would not exist. Individuals who agree to friend a researcher on Facebook are hypothesized to thus be more likely to heed get-out-the-vote messages embedded in that researcher's status updates, despite the lack of an offline friendship or close relationship. In contrast to a 2012 study, where turnout only increased among those exposed to banner ads featuring close Facebook friends, status updates from a non-close Facebook friend are hypothesized to be interpreted as coming from a trusted source and thus more effective at generating voter turnout because the targeted individual accepted the researcher's Facebook friend request.[3] A similar previous effort also using paid advertisements on Facebook also failed to increase voter turnout.[4]

This theory is tested with a **convenience sample** of university students invited to become Facebook friends with one of the authors. A convenience sample is exactly what it sounds like – a collection of participants in the study chosen in part because they are conveniently available to the researcher. Major advantages of convenience samples are that they are usually affordable – a key concern for many researchers – and that there are usually minimal hoops to jump through to use them. For example, many researchers do experiments with students at their home institution or on members of their local communities. The major disadvantage of convenience samples is that sometimes we can't be sure if the findings are generalizable to the broader population, for example if findings from a convenience sample of students would also apply to non-student populations.

The major hypothesis is that students **randomly selected** to be exposed to user-generated political content on Facebook will vote at a greater rate than will students randomly selected to be exposed to user-generated apolitical content. The power of the message is emphasized by the ongoing nature of the postings, creating a stronger feeling of closeness with the friend sharing those status updates. In other words, the power of assignment to the **treatment group**, or the group receiving the manipulation, is expected to be stronger than that observed by Bond *et al.* or Collins *et al.* both because the treatment is delivered via friend status updates rather than via a paid advertisement and because the treatment is repeated over time.[5] Variation in this feeling of closeness should also lead to heterogeneous effects: individuals in the treatment group who feel closer to the author because they share racial and gender identities (the posting author is a white woman) will exhibit larger increases in turnout. Assignment to the treatment group is expected to generate larger increases in turnout among younger students, defined here as those under the age of 30, both because younger voters are less likely to have rigid voting habits and because the posting author is a young person. In addition, because younger citizens are less likely to vote than older citizens, there is simply more room for improvement among younger voters. To summarize, the hypotheses of this study are:

H_1: Students randomly selected to be exposed to a series of political Facebook status updates from a Facebook friend will be more likely to vote than students randomly selected to be exposed to a series of

apolitical Facebook status updates from the same Facebook friend.

H_2: This effect will be stronger among individuals who are young (less than 30 years old).

H_3: This effect will be stronger among individuals who are white women (who share gender and racial identities with the Facebook friend making the postings).

The Research Strategy: What is a Randomized Field Experiment?

Randomized **experiments** have experienced a renaissance in the last two decades and are considered the gold standard for impact assessment and program evaluation.[6] What we mean here when we speak of randomized experiments is perhaps best understood as parallel to medical drug trials. In the medical field, controlled trials are often used to test new drugs or procedures. Individuals are recruited for a test and randomly divided into **treatment** and **control groups**. Those in the treatment group are administered the new drug or procedure, while those in the control group receive a placebo such as sugar pills or sham surgeries. Subsequent differences in health outcomes between individuals in the two groups can be clearly attributed to the treatment.

Randomized experiments can take many forms. A researcher might conduct a **pre-test/post-test design**, where the dependent variable of interest is measured both before and after participants are exposed to the treatment. For example, if you were interested in whether watching a short video about transgender people (compared to watching a short video about something unrelated, such as recycling) would change attitudes about transgender rights, you would survey both the control and treatment groups to measure their attitudes about transgender rights both before the groups watched the videos and afterwards. This classic design is illustrated below.

Experiments with a **post-test only design**, like the one described here, don't

include a measure of the dependent variable before exposure to the treatment. This is often used when either there is no relevant pre-test measure (you can't measure whether people have voted before Election Day) or if pre-testing is simply not possible or prohibitively expensive.

	Pre-test	Manipulation	Post-test
Experimental Group	O_1	X	O_3
Control Group	O_2		O_4

Figure 11.1 A Classic Experimental Design

Another common experimental design is a **factorial design**, where more than one variable is manipulated. For example, you might test the power of a get-out-the-vote message as well as the timing of delivery, comparing the effect of the same message delivered a week before Election Day to one delivered the evening before Election Day. There are actually quite a number of ways in which a randomized experiment can be designed and implemented; researchers can customize their plans to help them answer their research questions appropriately.

Similar to medical drug trials, experiments in political science *randomly assign* individuals to receive a treatment expected to generate a change in attitude or behavior, while others are randomly assigned to receive a placebo treatment or to no treatment whatsoever. Subsequent observed differences in attitudes or behavior can then be clearly attributed to the treatment message. For example, an experiment seeking to increase voter turnout might expose some individuals to a postcard or telephone blandishment to vote, while those in the control group would receive either no message or one encouraging recycling (a message not expected to affect voter turnout). **Random assignment**, where all participants have an equal chance to be placed in the control or treatment groups, helps to assure that any difference between the two groups is the result of the treatment, not some bias in the group. While survey research can only show **correlation** between an independent and dependent variable, experiments give us greater confidence in the cause and effect relationship between variables. Hundreds of randomized experiments of this sort have been conducted over the last decade, generating valuable insights in a variety of social science disciplines.

Issues with this Research Strategy

Field experiments are studies conducted in real-world settings, as opposed to in a research lab. They are particularly valuable because they provide evidence of causality (**internal validity**, or confidence that the treatment caused the outcome), as well as robust **external validity** (confidence that the findings can be generalized to other situations and hold up outside the lab). For example, as noted above, a convenience sample of undergraduate students might not provide findings that can be generalized to the adult non-student population. Studies using participants that are all men or all of one racial or ethnic group might not be generalizable to women or to members of other ethno-racial groups.

While observational studies have produced tremendous advances to our understanding of how people think and act in the political arena, randomized experiments allow for robust hypothesis-testing that pinpoints causal mechanisms. Observational studies of voter mobilization is particularly tricky because candidates and campaigns often target likely voters in their mobilization efforts, voters often have inaccurate recall of whether or not they were mobilized, and because likely voters are more easily contacted by those seeking to encourage participation. This makes field experiments particularly valuable for research exploring how best to increase voter turnout, because through random selection of participants into treatment and control groups it holds constant other potential predictors of the vote, and because validated turnout data can be obtained from local political offices.

On the other hand, field experiments tend to be expensive and difficult to implement. For example, door-to-door experiments using participants encountered in their homes requires transporting research staff, and ensuring their safety, and targeted individuals are often not at home, requiring extensive time in the field and numerous contact attempts. Experiments that use mailed materials require funds for printing and postage. Experiments that target groups that use multiple languages require either hiring research staff fluent in multiple languages, or paying to have written materials translated. These and other financial and logistical challenges of field experiments often lead researchers to instead opt to conduct laboratory experiments.

Ethics are a further constraint on field experiments. Not everything can

ethically be randomized. For example, you should not withhold medical treatment from some folks infected with a virus, to see how the disease will progress in the absence of treatment. This is what happened with the infamous Tuskegee Syphilis Study, where some Black Americans with syphilis were deliberately left untreated. You should not deliberately lie to voters to see how the information given to them affects their attitudes or behavior. For example, you should not falsely tell some voters that one of the candidates is corrupt or has had an extramarital affair in order to see whether that affects their feelings of political trust or their vote choice. Federal regulations (and common sense) also restrict experiments conducted on certain vulnerable populations, such as children or prisoners.

The Literature Review: Previous Research on Voter Mobilization

In order to begin our research, we had to first conduct a **literature review**. A literature review, an investigation of what prior studies have found in the area of study, helps us to understand what is already known and what areas need further exploration. A literature review helps us develop reasonable hypotheses, choose an appropriate and fruitful research method (considering the strengths and weaknesses of prior methods used), and build on existing knowledge. We know from our review of the scholarly literature that both direct (e.g., face-to-face conversations and live telephone calls) and indirect methods (e.g., text messages or mailed postcards) of encouraging targeted individuals to vote deliver the same information: that an election is approaching, and that the targeted individual should vote. Yet, simple exposure to the information is often insufficient to trigger compliance.

A recent **meta-analysis** published in 2013 by Green, Aronow, and McGrath of both published and unpublished work concludes that generating larger effects requires either personal methods or the inclusion of strong social pressure messages.[7] A meta-analysis brings together findings from multiple experiments or studies in a combined analysis. Combining data in this way means a larger

sample and more precision in estimated effects. Green, Aronow, and McGrath note: "When social norms are asserted forcefully, the effects tend to be quite large, and even pre-recorded phone calls conveying social pressure messages significantly increase turnout."[8] Another set of experiments using impersonal methods suggests that the source of the GOTV message can have important consequences. In 2007, Nickerson found that nonpartisan emails from a variety of nonpartisan organizations did not increase turnout, but in a series of experiments in 2012 Malhotra, Michelson, and Valenzuela found that email messages from the local registrar of voters did increase turnout, and unpublished work suggests that email messages from personal acquaintances are even more powerful.[9] Bond et al.'s 2012 Facebook experiment found effects among close friends, but not among ordinary Facebook friends.[10]

A final related trend in GOTV research is that target populations are increasingly difficult to reach using traditional methods, necessitating a shift in communication methods.[11] Door-to-door efforts, while potentially very effective, are time consuming and expensive, in part because targeted voters are not always at home waiting to be contacted. Telephone canvassing is more cost-efficient, but targets are increasingly unwilling to answer calls from unknown or unidentified numbers, making strong contact rates, and thus treatment effects, elusive. Mail delivered via the US Postal Service is increasingly the realm of junk mail. Instead, individuals are communicating via text messages and social networking sites, including Facebook, Twitter, Instagram, and other new platforms. In order to get out the vote, messages need to be delivered to targeted voters where they are: online.

Social networking generates social capital and political participation. In a 2007 survey of 286 undergraduates at Michigan State University, Ellison, Steinfield, and Lampe find that intensity of Facebook use is strongly related to social capital.[12] In a 2009 web survey of 2,600 Texas college students, Valenzuela, Park, and Lee find small but statistically significant relationships between Facebook use and social capital, civic engagement, and political participation.[13] Bode finds that particular Facebook activities generate various types of political participation.[14] Pasek, More, and Romer find that use of social networking sites such as Facebook can generate a culture that encourages social capital.[15] Using the Pew Internet & American Life Project's Spring Tracking Survey from 2008, Gainous, Marlowe,

and Wagner find a strong positive relationship between online social networking and online political participation, which they attribute to the building of social capital.[16] Looking at the Pew data and a sample of students from the University of Louisville and Florida Atlantic University, Gainous and Wagner find in both datasets that heightened Internet social networking predicts participation.[17] In sum, there is a growing body of evidence, based on survey data, that social networking sites such as Facebook generate social capital and political participation.

Gainous *et al.* further hypothesize that social networking will increase political participation when that networking includes political exchanges.[18] They find that "for every 1 unit increase in political SNS use, there is a 0.33 unit increase in online political participation."[19] Bond *et al.* also find evidence of the effect of political messages on participation.[20] Using a randomized field experiment with 61 million Facebook users, they find that exposure to a banner advertisement encouraging voting in the November 2010 election and noting participation by one's Facebook friends increased turnout by 0.39 percentage points, but the effect was limited to those whose messages were from close friends.

At the time of this study, American Internet users spent 53.5 minutes a month on Facebook – more than any other website.[21] Facebook members use the site for a variety of reasons, but among the most common is to keep up with friends. Individuals are more likely to comment on or like posts from their friends than to post their own news, and Facebook use leads to increased awareness of others' actions.[22] Thus, individuals who use such sites are more likely to know if others in their network are politically interested and active. This study tests the effect of that knowledge on subsequent behavior. Specifically, it tests the hypothesis that randomly assigned exposure to political SNS messages that encourage voting, operationalized as Facebook status messages, will increase the likelihood of an individual voting.

Conducting the Study: Collecting the Data and doing the Analysis

In the fall of 2010, one of the authors visited undergraduate political science classes being taught at a large southern public university. The visited classes were sections of an introductory political science course required of all students at the university as part of the core curriculum; thus, the recruited students represented a sample of all students at the university, not only political science majors or students particularly interested in politics. The visiting author, a graduate student at the university, identified herself as a fellow student working on a research project for a class in which she needed to observe how people use Facebook to talk about politics. The participants were in no way familiar with or connected to the authors in or out of the classroom prior to the study; the graduate student whom they were asked to friend was not their teaching assistant or known to the subjects in any other way. Participants were recruited in person, but did not have an existing offline relationship with the authors and were not personally known to the authors prior to the launch of the experiment. At no other time during the experiment did the authors have contact with the students outside of Facebook.

To participate in the project, students had to be enrolled in one of the political science classes at which the recruitment request was made, have an active Facebook account at the time of enrollment, send the author a friend request before the deadline that included their name and full birth date, and be at least 18 years of age on Election Day 2010. Approximately 2,800 students were invited to participate; 604 students enrolled in the experiment by friending the author on Facebook.

Informed consent was obtained from all participants during the enrollment process, consistent with standard **Institutional Review Board (IRB)** procedures. In response to a number of ethically questionable research studies in the early twentieth century, including the Tuskegee Syphilis Study, researchers who use human (and animal) subjects are now required by universities to get prior approval from Institutional Review Boards. The certification and review of one's study required to get IRB approval helps assure that participants in a study are treated ethically and that they have given **informed consent** to participate and are aware of the benefits and potential risks of their participation as well as of their right to withdraw from the study at any time. Participating students knew that they were part of a study, although they were not told the specific hypotheses or dependent variables that would be measured. The consent form

noted, "The purpose of [this] study is to understand the link between social networking sites and politics." They were also informed that the "researcher may also use your name and/or the information you provide by enrolling in the study to match public records, such as electoral participation, to study participants." Students were not told directly that the study was designed to influence turnout.

Given that they signed these informed consent forms at the outset of the experiment, a possible concern is that participants changed their behavior and were more likely to vote because they knew they had enrolled in a study about political behavior and they were being watched. Such **experimenter effects**, also sometimes called guinea pig effects or the **Hawthorne effect** after a study where participants' behavior seemed to have change as a result of being studied, are likely negligible. First, research has shown that very few individuals read the fine print of similar agreements, such as end user license agreements (EULAs) required to join social networking sites or when making a purchase from an online retailer. For example, Bakos, Marotta-Wurgler, and Trossen find that fewer than 0.1 percent of retail software shoppers choose to access the EULA, and that even fewer read more than a small proportion of those agreements, evidenced by the median time spent viewing the EULAs of just 29 seconds.[23] Similarly, 91 percent of Facebook users have never read the site's Terms of Service and 89 percent have never read Facebook's Privacy Policy.[24] The effect of signing the consent form likely had negligible effects on subsequent behavior: individuals assigned to both the treatment and control groups viewed and signed the same consent form, yet their subsequent turnout behavior differed dramatically. At the same time, it is possible that individuals randomly assigned to the treatment group were reminded by the author's political Facebook posts that their political behavior was being monitored, and thus voted in order to please the researcher. Information collected from participants' Facebook pages during the experiment suggests this is not the case. We compared individuals randomly assigned to the treatment group to individuals randomly assigned to the control group and found that there was no systematic difference in the likelihood of participants claiming on their Facebook page that they had voted (22.02% vs. 18.75%) or of posting political information of their own (30.06% vs. 27.38%). Neither of these differences is statistically significant. In other words, the public (Facebook) political behavior of both groups was indistinguishable, but the

private (voting) information – which participants had technically been notified in the consent form might be consulted – shows a clear treatment effect, as detailed below.

Among the 604 enrolled participants, 304 were randomly assigned to the treatment condition and 300 were randomly assigned to the control condition. Because Facebook users can choose to opt out of receiving information from individual friends, participating students were told that the postings would include two extra credit exam questions. Students unable or unwilling to participate could email the author to have the questions sent to them when they were posted to Facebook. Over the course of the experiment 51 students sent such requests. This incentive was designed to limit the extent of the **failure-to-treat problem**, wherein individuals randomly assigned to receive a treatment in a field experiment are not exposed to the treatment. Exposure to the postings by the author – for both the treatment and control groups – depends on those postings appearing on the students' News Feeds. Yet, there is no mechanism for measuring such exposure, and thus actual treatment effects cannot be estimated. Instead, the results are limited to **intent-to-treat effects**. Intent-to-treat means precisely what it sounds like: this is the outcome among folks that we intended to treat. In some experiments, participants targeted to receive a treatment may not actually be treated. They may delete the treatment email without reading it, throw away the treatment mailer without opening the envelope, or just never answer the telephone or the door when the researchers are trying to contact them. While researchers know when someone was never contacted by phone or at the door, it's harder to know if they're reading their mail or listening to the radio. Examining intent-to-treat effects is appropriate when the researcher can't know whether participants actually received the intended treatment. Based on prior research on indirect methods such as television and radio advertisements, email and text messages, we feel confident that the number of students who chose to block the experimental status update messages is likely to be quite small. Similar research found that at the time of this study only 3 percent of participants hid the author's postings from their Facebook page.[25]

Once they agreed to include the author in their online social network, students were exposed to messages with varying degrees of political content. Students in the treatment group were exposed not only to information about the upcoming

election, such as when to vote, but also to the norm of participation by the author with whom they had become Facebook friends. The Friends List feature on Facebook was used to control which participants were exposed to specific pieces of content. The Friends List feature provides users with a way to categorize their friends into subsets or lists. For this experiment, participants were randomly assigned to lists titled Treatment and Control, allowing for manipulation of the audience of a specific status update while ensuring that all of the other content available from the author's profile was identical between conditions. Facebook did not allow users to see how other users had categorized them, so participants were unaware of their condition assignment. In addition, no participants had access to the author's political affiliation, relationship status or work history via Facebook.

All participants received a total of 23 status updates from the author: three stimuli the first week, four stimuli the second week, seven stimuli the third and fourth weeks, and two stimuli the fifth week – the day before Election Day and Election Day. Members of each list received their stimuli on the same day and at about the same time. Students in both groups were exposed to nine apolitical stimuli, with the intent of making the author seem more like a real person and not one solely interested in the election. Thus, the author was embedded into individuals' online networks. For the remaining 14 occasions, the control condition continued to receive apolitical stimuli, while the treatment condition received messages that encouraged them to think about the election and provided logistical information, such as the date of the election or poll times for voting in the November 2010 election. On Election Day, the status update sent to students in the treatment condition was an explicit reminder to vote.

That those assigned to the treatment and control groups felt the author was a member of their community was illustrated by their tendency to like or comment on the author's postings. Rather than just passively viewing the information being posted, participants engaged in conversation with the author. Members of the treatment and control groups were equally likely to comment on or like the non-political postings to which both were exposed: over the course of the experiment there were 29 likes or comments to the author's postings, including 16 from the treatment group and 13 from the control group. Yet, the author was clearly not a true or close friend. They met her only once, at the classroom

session when they were invited to participate in the experiment, and had no further offline contact with her.

At the time of this experiment, Facebook's terms of service asked users not to collect other users' content or information using automated means. To not violate this policy, individual screen shots were taken of participants' profile pages and the information contained on those pages was hand coded. After the election, each participant was searched for in the list of registered voters from the state voter file, using their birth date and other data collected from their Facebook profiles. Through this process, 59 percent of the participants assigned to the control condition were matched to the voter file compared to 55 percent of participants in the treatment condition. This difference is not statistically significant – exposure to the norm of political participation did not cause students randomly assigned to the treatment group to register to vote. However, the experiment did not launch until it was too late to register and participate in the November 2010 election. Moreover, the status updates posted to the treatment group did not include an explicit reminder to register or instructions on how to do so; it is possible that exposure to a conversation where the social norm of registration, rather than turnout, was modeled, would be more effective at increasing voter registration, particularly if such an effort were conducted prior to the close of registration for the target election.

A total of 344 participants could be positively matched to the state voter file, 51 percent from the control condition (N=176) and 49 percent from the treatment condition (N=168). The difference is not statistically significant. Focusing on those 344 participants who were matched to the state voter file, the treatment and control groups are balanced. The randomization produced the expected similarity on observable characteristics between the treatment and control groups (see Table 11.1).

Table 11.1 Percent Voting, by Assignment to Treatment Group (Ns and standard errors in parentheses)

ALL (N=344)		
% Voting, control group	% Voting, treatment group	Intent-to-treat effect
22.73	30.95	8.22[*]−
(40/176)	(52/168)	(4.77)
Age<30(N=318)		

18.5	29.5	10.97[**]
(30/162)	(46/156)	(4.76)

White Women (N=135)		
13.24	32.84	19.6[**]
(9/68)	(22/67)	(7.09)

Source: Calculated by authors using data collected from participants' Facebook pages and Georgia state voter file.

Notes: [**] = p<.01, [*] = p<.05, one-tailed tests.

Differences between control and treatment across a host of demographic and political variables were both substantively small and statistically insignificant (see Table 11.2). The one exception to this is the average number of Facebook friends: those assigned to the control group have more friends. In the analysis below, several steps are taken to correct for this imbalance. First, number of Facebook friends is included as a covariate in the **multivariate logistic regression models** (more about this method will be discussed below). Second, one of the models is run without outliers, defined as individuals with more than 1,675 or fewer than 50 friends, which eliminates the imbalance (see Table 11.3).

Table 11.2 Summary Statistics and Balance Tests

	Control Group	Treatment Group	P-value
Mean age	21.3	21.7	0.43
% male	32.0	29.2	0.57
% Black	40.5	40.1	0.95
Mean # of Facebook friends	600.1	505.8	0.03
Mean # of mutual Facebook friends	4.5	3.6	0.14
% voted in 2008	**2.8**	3.3	0.27

Source: Calculated by authors using data collected from participants' Facebook pages and Georgia state voter file.

Note: *p*-values are two-sided. See Table 11.3 for an alternative version of the randomization check using logistic regression.

As can be seen in Table 11.1, about 31 percent of participants randomly assigned to the treatment condition voted on November 2, 2010 compared to 22.7 percent of participants assigned to the control condition, a massive and **statistically significant** difference of 8.22 percentage points (S.E. = 4.77) based on a difference in proportions test. Statistical significance, represented by the "p," or **p-value**, in the table tell us what the "p"robability is that we would find these means in the population randomly if in fact there was no relationship between the independent and dependent variables. Generally speaking, scholars say that a p-value of less than .05, which means there is less than a 5 percent chance one would have found these differences randomly, gives us a pretty high level of confidence that we have found some meaningful differences between the groups.

The **effect size** – the estimated effect of random assignment to the treatment group – increases to 10.76 percentage points (S.E. = 4.76) when restricting the analysis to those younger than 30 years old (see Table 11.1). This effect size is comparable to that found using previous GOTV tactics, including door-to-door canvassing, two-round phone banks, and the strongest social pressure direct mailings used in the Gerber, Green, and Larimer effort.[26] Because the friending author was a white woman, white women in the experiment were expected to be even more likely to consider her as part of their in-group, and thus be the most likely influenced to vote. This hypothesis is also supported by the turnout data: among white women, the effect size is 19.6 percentage-points (S.E. = 7.09).

Because the dependent variable (voting) is dichotomous, the data are also analyzed using **logistic regression** (see Table 11.3).

Table 11.3 Logistic Regression: Effect of Assignment to Treatment Group on Turnout

	Coefficient Estimate	Standard Error	P-value
Model A: ALL (N=344)			
Treatment	.421±	.245	.086
Constant	−1.224**	.180	.000
MODEL B: ALL, with covariates (N=344)			
Treatment	.445±	.266	.094
Age	.023**	.036	.007

Age	.092—	.034	.007
Male	.515±	.278	.063
Black	.273	.272	.315
Number of FB friends	−.00004	−.0005	.928
Number of friends who voted in 2010	.076	.167	.651
Mutual friends	.027	.037	.464
Voted in November 2008	.951**—	.313	.002
Constant	−4.024**—	.855	.000

MODEL C: UNDERAGE 30 (N=318)

Treatment	.601*—	.283	.034
Age	.038	.100	.700
Male	.476	.295	.107
Black	.340	.285	.233
Number of FB friends	.0001	.0005	.811
Number of friends who voted in 2010	.055	.164	.740
Mutual friends	.021	.037	.557
Voted in November 2008	1.050**—	.373	.005
Constant	−3.102	2.059	.132

MODEL D: NO OUTLIERS N=294

Treatment	.504±	.285	.077
Age	.105**—	.038	.005
Male	.517±	.298	.083
Black	.170	.292	.560
Number of FB friends	.0001	.001	.893
Number of friends who voted in 2010	.084	.168	.618
Mutual friends	.005	.040	.905
Voted in November 2008	.905**—	.338	.007
Constant	−4.133**—	.936	.000

MODEL E: WHITE WOMEN (N=135)

Treatment	1.093*—	.482	.023
Age	.094	.058	.103

Number of FB friends	.0003	.001	.724
Number of friends who voted in 2010	.120	.274	.662
Mutual friends	−.085	.086	.322
Voted in November 2008	1.042*	.524	.047
Constant	−4.245**	1.437	.003

MODEL F: WHITE MEN (N=62)

Treatment	−1.034	.718	.150
Age	.103	.085	.226
Number of FB friends	−.001	.002	.486
Number of friends who voted in 2010	.156	.475	.743
Mutual friends	.133	.106	.211
Voted in November 2008	2.090*	.871	.017
Constant	−3.371	2.088	.107

MODEL G: BLACK WOMEN (N=95)

Treatment	.110	.499	.825
Age	.032	.082	.695
Number of FB friends	.001	.001	.462
Number of friends who voted in 2010	−.062	.313	.844
Mutual friends	.041	.059	.492
Voted in November 2008	1.050±	.621	.091
Constant	−2.672	1.897	.159

MODEL H: BLACK MEN (N=38)

Treatment	1.959*	.919	.033
Age	.277±	.156	.075
Number of FB friends	.0004	.002	.777
Number of friends who voted in 2010	.708	.723	.328
Mutual friends	.003	.128	.983
Voted in November 2008	−.140	.983	.887
Constant	−7.699*	3.769	.041

± = p < .10, * = p<.05, ** = p<.01, two-tailed. In order to avoid dropping observations in models with covariates, missing data is imputed using the Amelia II package in R, version 1.7.2 (Honaker, King and

Blackwell 2013). The one exception to this is Model D (no outliers), which was run with the original data instead of the imputed data.

Source: Calculated by authors using data collected from participants' Facebook pages and Georgia state voter file.

A dichotomous dependent variable has only two values, such as yes or no, or voted or didn't vote. The coefficient estimates from a logistic regression can be somewhat difficult to interpret, however, because they are predicting the log odds of an event – the probability that the dependent variable will change from one value (did not vote) to the other (voted). You don't need to understand how to calculate log odds to understand the results from this experiment. Just keep in mind that a positive estimate for the independent variable (assignment to treatment) means that the likelihood of voting is increased, while a negative estimate would mean that the likelihood of voting is decreased.

The multivariate models confirm that students randomly assigned to the treatment condition were more likely to vote than those randomly assigned to the control condition. As hypothesized, the effect is even stronger among youth, defined as those younger than age 30. The effect of assignment to the treatment group is particularly strong among white women and Black men, although these subsets of the data are relatively small; future studies with larger subgroup samples are needed to confirm these findings. Consistent with previous work, age and voter history are statistically significant predictors in most of the models as well, with older individuals and those who voted in 2008 more likely to participate in 2010.

In other words, messages about the election shared on Facebook did make students in this experiment more likely to vote. Participants weren't targeted personally, or tagged in any of the messages. They just were exposed on their normal Facebook feed to a series of posts about voting and the election. At the same time, the outreach was personal, because the posts came from an individual they had chosen to friend on Facebook. This sort of friend-to-friend get-out-the-vote model has also worked in other experiments. People are influenced by social norms – by what their friends are doing – including voting.

Lessons to Be Learned

Researching developing technology can be challenging. For example, Facebook is a private company that guards how it prioritizes content shown to its users. Therefore, researchers cannot be guaranteed that the process is the same over time or across users. This challenge is what made an experimental design such a good fit for studying the effect of exposure to social network site messages with varying degrees of political content. That said, creating this experiment had its own challenges. First, getting approval from the university to conduct the research was cumbersome because we had to be able to prove that we wouldn't be causing the participants any psychological distress by forging this manufactured "friendship." For this reason, the friending author has never "unfriended" anyone who participated in the study. Second, it is important to conduct your experiment within the terms of service a company like Facebook defines. In our case, Facebook prohibits scraping activity of users' profiles. This means that all the data collected from the participants' profiles that were used to match them to the voter file or conduct additional analysis needed to be collected manually. Since coding each profile in real-time would take too long, we enlisted helpers to take screenshots of each participant's profile that were then coded at a later time. Finally, finding participants for this study was difficult. The original design for this study included a non-student sample. One of the authors went to several randomly selected grocery stores in her neighborhood and handed out flyers asking strangers to friend her on Facebook. She handed out over 1,000 flyers and received just 25 friends. The student sample was easier to motivate since we could provide them with an incentive.

In this study we operationalized "exposure" as Facebook friend status updates, in an attempt to create a positive impact on the political participation of students at a large southern public university. The treatment generated mobilization effects parallel to those from the most forceful social pressure and direct method techniques using an extremely cost-effective social networking platform. The only cost incurred was the time to recruit friends and make the Facebook postings.

Unfortunately, however, given the relative ease and low cost of sharing

political content through social media and the effectiveness of this content impacting voter behavior, validated peer-to-peer social media influence could be used in a negative context – to suppress voter turnout or sway individuals' opinions about candidates by creating content containing misinformation. In fact, Facebook has recently expanded its Elections Integrity department to help deter the platform from being used to negatively impact the democratic process. While Facebook's motives appear genuine and well intentioned we have to wonder what, if any, impact on democracy will result from having a private company like Facebook censoring content posted to its site in the name of preserving democracy.

Clearly, additional experiments are needed to test if similar effects can be generated using diverse populations, different messages, such as partisan and persuasive messages, or with broader geographical scope. Manipulations regarding the number of posts, timing of posts, and type of posts, such as hyperlinks, status updates, direct appeals or informational posts are needed to better understand the extent to which social media appeals can influence the electorate. However, in addition to our thoughts as researchers, findings like those presented here can draw attention to issues that society more broadly needs to address philosophically. In this case, how do we, as an electorate, feel about being manipulated – either positively or negatively? Do users of sites owned by private companies have a right to know how the content being presented to them is being prioritized? To what extent are we comfortable with the private companies who own sites like Facebook taking on the responsibility of policing content posted by its users? Is it in the best interest of social media companies to censor content?

While we think about how social media users and the companies that provide those services act, it is important to remember that not all individuals are Facebook users or are willing to be friended online – there is an opt-in component here that may screen out those for whom this sort of GOTV messaging would not be effective. At the same time, use of online social networking sites is widespread and growing. At the time of this study Facebook had a billion users worldwide and 95 million in the US, and American users spend more time on Facebook than on any other website. As of 2017, Facebook has doubled their worldwide reach to two billion users.

The increasing dominance of online networks suggests such efforts would likely be successful and could have large effects on turnout among the increasingly wired US population. The ease of incorporating new people into one's online network means that even those who are minimally known to each other become online friends, strengthening weak ties and creating new ones. Young people tend to be nonvoters, and turnout is generally low for midterm elections; yet, in this experiment turnout among students for a midterm election was increased by 8.2 percentage points, and among younger students by nearly 11 percentage points.

Hundreds of GOTV experiments conducted in the last decade conclude that, absent social pressure or a personal sociocultural interaction, reminding registered voters about an upcoming election is unlikely to move many voters to the polls. Here, in contrast, political Facebook status updates generated large increases in turnout. In recent years, individuals have increasingly moved from face-to-face social networking to participation in online social networking sites. In order to increase voter turnout, scholars and practitioners need to go where people are congregating – on Facebook, Twitter, LinkedIn and other social networking sites. As Gainous *et al.* note: "a shift to an online world is likely to magnify the magnitude and significance of SNSs over time." Today's millennial generation – college students – represent that future, and digital GOTV messages are the future of political mobilization.[27]

Interested to Know More about the Study Discussed in this Chapter?

Consult the research publication:

Teresi, Holly, and Melissa R. Michelson. 2015. "Wired to Mobilize: The Effect of Social Networking Messages on Voter Turnout." *Social Science Journal* 52(2) (June): 195–204.

Exercises and Discussion Questions:

1. What do you see as the primary strengths of field experiments? What are their primary weaknesses?
2. Identify a research project discussed in this textbook that does not use a field experiment and develop a proposal for one which would provide further insight into the research question being asked.
3. Is it ethical for researchers to conduct experiments on the public? Should they always be told that they are being observed? How might informing folks that they are being observed complicate a field experiment on voter turnout or other socially desirable behavior?
4. Is it ethical to manipulate turnout, given the potential for the research to have an impact on election results?

Recommended Resources:

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: Norton.

Time-Sharing Experiments for the Social Sciences (http://tess.experiments.org): This website is the location for a project funded by the National Science Foundation to allow scholars to conduct experiments that include a very large, diverse sample, thus allowing for greater certainty of the generalizability of one's findings. The experiments are conducted through the Internet on a random sample of the population and include "time-sharing" whereby different research questions posed by different scholars are considered in the same experiment. Scholars interested in using this service must submit a proposal for consideration. You can access the projects that have been conducted and the data that has been collected to examine and analyze for yourself.

Notes

1 Green and Gerber 2015; García Bedolla and Michelson 2012.

2 Kobayashi 2010.

3 Bond *et al.* 2012.

4 Collins, Keane, and Kalla 2014.

5 Bond *et al.* 2012; Collins, Kalla, and Keane 2014; Michelson, García Bedolla, and McConnell 2009.

6 Green and Gerber 2015.

7 Green, Aronow, and McGrath 2013.

8 Green, Aronow, and McGrath 2013: 36.

9 Nickerson 2007; Malhotra, Michelson, and Valenzuela 2012; Davenport 2008.

10 Bond *et al.* 2012.

11 Green and Gerber 2015; Issenberg 2012.

12 Ellison, Steinfield, and Lampe 2007.

13 Valenzuela, Park, and Kee 2009.

14 Bode 2012.

15 Pasek, More, and Romer 2009.

16 Gainous, Marlowe, and Wagner 2013.

17 Gainous and Wagner 2011.

18 Gainous, Marlowe, and Wagner 2013.

19 Ibid: 154.

20 Bond *et al.* 2012.

21 Hampton *et al.* 2011; Nielsen 2012.

22 Joinson 2008.

23 Bakos, Marotta-Wurgler, and Trossen 2009.

24 Jones and Soltren 2005.

25 Teresi 2009.

26 Gerber, Green, and Larimer 2008.

27 Gainous, Marlowe, and Wagner 2013: 155.

References

Bakos, Yannis, Florencia Marotta-Wurgler, and David R. Trossen. 2009. "Does Anyone Read the Fine Print? Testing a Law and Economics Approach to Standard Form Contracts." NYU Center for Law, Economics and Organization Working Paper No 09-40. http://ssrn.com/abstract=1443256.

Bode, Leticia. 2012. "Facebooking It to the Polls: A Study in Online Social Networking and Political Behavior." *Journal of Information Technology & Politics* 9(4): 352–369.

Bond, Robert M., Christopher J. Fariss, Jason H. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 498: 295–298.

Collins, Kevin, Laura Keane and Josh Kalla. 2014. "Youth Voter Mobilization Through Online Advertising: Evidence from Two GOTV Field Experiments." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, Illinois.

Davenport, Tiffany C. 2008. "Unsubscribe: Comparing the Effects of Peer-to-Peer and Mass Email Messages on Voter Turnout." Poster presented at the annual meeting of the American Political Science Association, Boston, MA.

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12(4): 1143–1168.

Gainous, James, and Kevin M. Wagner. 2011. *Rebooting American Politics: The Internet Revolution.* Lanham, MD: Rowman and Littlefield.

Gainous, Jason, Adam D. Marlowe, and Kevin M. Wagner. 2013. "Traditional Cleavages or a New World: Does Online Social Networking Bridge the Political Participation Divide?" *International Journal of Politics, Culture, and*

Society 26(2): 145–158.

García Bedolla, Lisa, and Melissa R. Michelson. 2012. *Mobilizing Inclusion: Transforming the Electorate Through Get-Out-the-Vote Campaigns.* New Haven, CT: Yale University Press.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(1): 33–48.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton & Company.

Green, Donald P., and Alan S. Gerber. 2008. *Get Out the Vote: How to Increase Voter Turnout*, 2nd edn. Washington, DC: Brookings Institution.

Green, Donald P., Peter M. Aronow, and Mary C. McGrath. 2013. "Field Experiments and the Study of Voter Turnout." *Journal of Elections, Public Opinion and Parties* 23(1): 27–48.

Hampton, Keith N., Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. "Social Networking Sites and Our Lives: How People's Trust, Personal Relationships, and Civic and Political Involvement Are Connected to Their Use of Social Networking Sites and Other Technologies." *Pew Research Center's Internet & American Life Project*, http://pewinternet.org/Reports/2011/Technology-and-social-networks.aspx.

Honaker, James, Gary King, and Matthew Blackwell. 2013. "Amelia II: A Program for Missing Data." Version 1.7.2, http://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf.

Issenberg, Sasha. 2012. *The Victory Lab: The Secret Science of Winning Campaigns.* New York, NY: Crown Publishers.

Joinson, Adam N. 2008. "Looking At, Looking Up or Keeping Up With People? Motives and Use of Facebook." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1027–1036. ACM.

Jones, Harvey, and José Hiram Soltren. 2005. "Facebook: Threats to Privacy." *Ethics and the Law on the Electronic Frontier Course*, Massachusetts Institute of Technology, http://ocw.ispros.com.bd/courses/electrical-engineering-and-computer-science/6-805-ethics-and-the-law-on-the-electronic-frontier-fall-2005/projects/facebook.pdf.

Kobayashi, Tetsuro. 2010. "Bridging Social Capital in Online Communities:

Heterogeneity and Social Tolerance of Online Game Players in Japan." *Human Communication Research* 36(4): 546–569.

Malhotra, Neil, Melissa R. Michelson, and Ali A. Valenzuela. 2012. "Research Note: Emails from Official Sources Can Increase Turnout." *Quarterly Journal of Political Science* 7(3): 321–332.

Michelson, Melissa R., Lisa García Bedolla, and Margaret A. McConnell. 2009. "Heeding the Call: The Effect of Targeted Two-Round Phonebanks on Voter Turnout." *Journal of Politics* 71(4): 1549–1563.

Nickerson, David W. 2007. "Does Email Boost Turnout?" *Quarterly Journal of Political Science* 2(4): 369–379.

Nielsen. 2012. "The State of the Media: The Social Media Report 2012." The Nielsen Company. www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2012-Reports/The-Social-Media-Report-2012.pdf.

Pasek, Josh, Eian More, and Daniel Romer. 2009. "Realizing the Social Internet? Online Social Networking Meets Offline Civic Engagement." *Journal of Information Technology & Politics* 6(3/4): 197–215.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community.* New York, NY: Simon & Schuster.

Teresi, Holly A. 2009. "Friending Your Way to Political Knowledge: A Field Experiment of Computer-Mediated Social Networks." Thesis, Georgia State University. http://scholarworks.gsu.edu/political_science_theses/29

Valenzuela, Sebastián, Namsu Park, and Kerk F. Kee. 2009. "Is There Social Capital in a Social Network Site? Facebook Use and College Students' Life Satisfaction, Trust, and Participation." *Journal of Computer-Mediated Communication* 14(4), 875–901.

Normative and Ethical Considerations of Political Science Research

Elizabeth S. Smith and Akan Malici

CONTENTS

In the 1960s, a sociology graduate student by the name of Laud Humphreys conducted a study of homosexual behavior in public places. Unbeknownst to the subjects of his field research study, Humphreys, a gay man himself, identified men having homosexual encounters in public bathrooms. He obtained the license plate numbers of these men and then went to their homes (disguising his own appearance so he would not be recognized) under the guise of conducting a "social health survey" and asked them questions about their personal life. He discovered that many of these men were in heterosexual marriages and living very traditional lives. In the era in which his research was done, the information

he gathered was remarkable as it fundamentally challenged notions of what was normal behavior and led to a deeper understanding of human sexuality. However, on many fronts, his study violated the most important ethical standards for research. Humphreys did not obtain informed consent from his subjects, deceived them regarding the purpose of his survey, and also infringed on their privacy by coming into their homes under false pretenses. Because of these ethical violations, he was denied his Ph.D. degree by his university.[1]

Because you are a college student, you may find at least equally disturbing the research conducted by Mary Henle and Marian Hubbell in the 1930s. These two psychologists focused on egocentrism in adult conversation (how often adults make reference to themselves as opposed to others when they speak). Specifically, they wanted to know whether there is a decrease in egocentrism as a person matures from childhood into adulthood. Their methodology was simple: They would carefully listen to conversations and count egocentric words or phrases in the conversations. Obtaining such relative frequency counts for their sample of young adults would allow them to compare their results to the results from samples of children's conversations. In order to be particularly unobtrusive, they did not shy away from secretly invading the privacy of the subjects being observed. Here is how they describe their procedure:

> In order not to introduce artifacts into the conversations, [we] took special precautions to keep the subjects ignorant of the fact that their remarks were being recorded. To this end [we] concealed [ourselves] under beds in students' rooms where tea parties were being held, eavesdropped in dormitory smoking-rooms and dormitory wash-rooms, and listened to telephone conversations.[2]

You may rightly be shocked, and it is good to know that this type of research would not be permissible any longer. Standards have changed over time and become more stringent to protect subjects participating in a research project. **Normative** and ethical questions and concerns must be considered at every stage of the research process in any field. Political scientists, like others, must consider the ethics of the research questions they ask as well as the methods by which they gather the answers to those questions. **Institutional Review Boards** (IRBs) are now present at all universities and any research using human or animal subjects (even research conducted by undergraduates) must be approved as

ethical by these boards before it is conducted.

Throughout the chapters of this book, we have seen the struggles and the joys surrounding the research process and particular research methods. Though this is the final chapter of this book, ethical and normative considerations should not be considered afterthoughts in this process. In fact, these considerations are at the forefront of any good scholarly work. So, what are the ethical and normative considerations any good scholar should take into account?

Normative considerations are value considerations concerned with how things (politics) should or ought to be. The Democratic Peace Theory that we discussed in [Chapter 2](), for example, carries strong normative implications. If democracies indeed do not fight each other, then the normative implication would be to want democratization in countries where it is lacking.

Ethical considerations are concerned with how one ought to act (as a researcher) morally speaking, with the rightness or wrongness of the researcher's actions, and the right and wrong ways to treat other living beings. The means that Henle and Hubel used to conduct their study on egocentrism in young adults just intuitively seem to be wrong and not ethical. The wrongness of our research activities may, however, not always be as obvious and so one must be very careful here.

Both normative and ethical considerations should matter to good scholars and they should be considered during every aspect of the research process: when one is choosing a research question and the appropriate or conducive methodology to study and answer it, when one is actually conducting the research, and finally when one is considering the meaning and implications of one's findings and conclusions.

Issues when Choosing One's Topic

Perhaps, the most important part of the research process is choosing an appropriate topic and research question. At the heart of this choice is considering the normative issue of whether this topic is important. When considering their **research question** and topic, good scholars ask themselves: Does this question

really matter? Does it address some fundamental problem or issue in our world? Does it contribute to making the world a better place? Does it challenge us to address in some way the fundamental question of how things ought to be in the world? Is it based on truthful premises and is it fruitful in suggesting answers to other problems? These are **normative questions**. While these questions seem lofty and the goals high, the ultimate purpose of any research project is to contribute in some way to understanding the world a little bit more in the hopes of making it a little bit (or a lot) better for all.

Certainly, the choice of topic depends in part on what you as an individual scholar think is important or significant or what should be considered as important by more people. However, as reviewed at the beginning of this book, excellent research is about revealing scientific knowledge, not promoting personal bias or opinion. While some research directly answers challenging questions, more often research contributes to just a portion of our understanding of the big picture. But, a good researcher should always be able to articulate an answer to the "so what?" question regarding the significance of their topic.

We believe that it is a good practice to use one's colleagues as bouncing boards for one's research ideas. As humans we are often guided by our own and very unique senses of curiosity. Led by our own curiosity, it often happens that one comes up with a particular (research) question and subsequently develops a blind passion for it, not seeing the limited usefulness or applicability of any answer to it. This is where colleagues can be helpful in giving us a reality check. Our colleagues, like ourselves, are experts on what we intend to study, particularly those who work in the same field as we do. They may be able to give a more objective answer to the "so what?" question.

Colleagues may also be helpful in another area. No matter how great our expertise in any given area may be, ultimately we are still limited. For example, even after conducting exhaustive **literature reviews**, it is quite likely that none of the scholars in this book have read every potentially relevant article or book on their topic. The questions we ask may have been asked and answered already, yet we may not be aware of it. Colleagues may save us a lot of time and obsolete engagement. If, however, our question is worth pursuing, colleagues may be helpful in suggesting literature that can in significant ways aid our research.

There are also other challenges when deciding on a research question.

Consider, for example, issues faced by scholars we met in this book when they were considering their particular topic of study. In Chapter 9, for example, authors Fleming, Cowen and Carlson took on a particularly contentious topic in their decision to research the school voucher issue. Passionate, partisan debate surrounds this issue with stakeholders like parents, the public and private school systems, the state and federal governments, and important political leaders having vested interests in certain outcomes. While this particular study was authorized by the state of Wisconsin, the legislature did not provide any funding to carry it out. These scholars then were forced to seek funds from various foundations and organizations. At the same time, however, they had to be sure they were not beholden to these groups to "find" certain outcomes or did not appear biased to outsiders because of these funding sources. This was an ethical challenge and it is a common problem when research is funded by private organizations which, of course, may have their own ideologies, interests and agendas and may want to see these supported by scholarly research, which rightly tends to be perceived as credible and authoritative. It is important for scholars to remain mindful of these challenges. Fleming, Cowen and Carlson explicitly drew our attention to this challenge. Yet, because they were mindful of it, they were confident that they could "handle" it while pursuing the very important question of whether these expensive and varied approaches to educating young people in a democracy were in fact efficacious.

The choice of topic for study also created particular ethical and normative dilemmas for two of the scholars who discussed their qualitative research in this book. In Chapter 4, Kate Kaup discusses her study of the treatment of ethnic minorities in China. The Chinese government, a normatively important entity for study precisely because it has been described by various commentators as a repressive regime, directly blocks access to information it deems politically sensitive, like research on human rights violations. Chinese scholars who dare examine such sensitive topics regularly find themselves imprisoned, as Ph.D. candidate Tohti Tunyaz discovered when he was sentenced to ten years in prison for his work on Uighur history.[3] It has also been in the experience of Western researchers who criticize the Chinese government that they can be denied visas to conduct future research. Given the often oppressive and, at times, dangerous research climate surrounding politically sensitive issues, the researcher may

understandably opt to pursue issues less critical of the government. But, according to Kaup and other scholars in similar positions, it is important that regimes like the Chinese government are not allowed to shape the research agenda as that would mean research is only done on sanitized political issues and presents a distorted view of these governments and their politics.

In [Chapter 5](), Kristina Thalhammer discusses her interviews with individuals who were subject to extreme regime violence and repression under the military junta in Argentina from 1976 through 1982. Her choice of topic was also clearly normatively important – when, how and why are people willing to resist unjust political authority? However, we should not fail to ask: Was it ethically responsible to go out and ask very sensitive and emotionally difficult questions of those who were involved? It is important to remember that loved ones of those asked were brutally tortured and/or were disappeared forever by the regime. Such a choice of topic required many participants to relive painful memories from their past or to confront their own lack of action in the face of injustice. A researcher must carefully weigh such costs against the potential benefits of pursuing a particular topic. The benefits of providing insight into political resistance and the fruitfulness of such a research pursuit for understanding other circumstances (like rescuing behavior during the Holocaust or government whistleblowers) were benefits that most would say outweighed the potential costs. Although she could not be entirely sure, it was Thalhammer's sincere assessment that her questions would not inflict enduring pain on the participants.

Some have argued that certain topics are off limits normatively speaking because they begin by suggesting premises that are biased or subjective or are not fruitful in providing us the potential to address other questions. Disagreements among scholars have been centered around (research) questions which ask, for example, "Did the Holocaust occur?" Such a question begins so far from a truthful premise, most would argue that it is illegitimate to ask. Or, for example, take the controversy surrounding the (research) question: "Are their genetic explanations for the differences between men and women in success in science?" Or, similarly, "Can we explain differences in educational achievement between blacks and whites based on innate traits?" Such topics are similarly considered by many normatively and ethically inappropriate in large part because they are reductionist about human behavior, suggesting that what we do and how we act

can be explained by simplistically examining categories assigned to us by birth rather than the totality of our experiences as humans and, in the asking, suggest no solution to the social problem.

There is clearly room for disagreement about the appropriateness of a choice of research question. A sincere and honest engagement with one's motivations for asking potentially controversial or offensive questions is a first step for serious scientific research as opposed to sensationalism or demagoguery. Once again we also find that it is smart to consult with one's colleagues. Talking to others, sharing one's ideas and taking suggestions and possible warnings seriously is always a good measure toward success as a reputable researcher.

Issues with One's Research Strategy

In 1971 a team of researchers led by Stanford University psychologist Philip Zimbardo conducted what came to be known as the (in)famous *Stanford Prison Experiment*.[4] The researchers were interested in studying abusive prison situations. They randomly selected two dozen students to live in a mock prison of the Stanford psychology building and they assigned to them randomly the roles of prisoners and guards. Zimbardo himself took the role of "Prison Superintendent." After a very short time (one day) participants adapted scarily well to their assigned roles. The guards displayed authoritarian attitudes and some went even so far as to display truly sadistic behaviors. As one form of punishment, for example, the guards would remove the prisoners' mattresses so that they would be made to sleep on the concrete floor. Equally shocking, some prisoners were forced to be nude as a form of punishment. At times, guards also did not allow the prisoners to urinate or defecate in a restroom, which contributed to deplorable sanitary conditions. The prisoners, on the other hand, adapted passive attitudes and endured physical abuse and, at the request of the guards, they went even so far to inflict punishment on their fellow inmates.

In the aftermath of the experiment, critics charged Zimbardo with losing sight of his true role as a psychologist as he permitted the abuse to continue for several days. This was an ethical charge. Indeed, it was only when a shocked graduate

student entered the situation and objected to the conditions of the mock prison that the experiment was terminated. At the time Zimbardo's study was cleared by the Ethics Code of the American Psychological Association. They did not anticipate that it would not only violate the dignity of the participants, but also cause them actual harm. It became evident that a research proposal on paper can look quite different from the actual conduct of the study. Moreover, it also became clear that the general "do no harm" principle established by Institutional Review Boards is not a guarantee for the research process to unfold accordingly.

When scholars choose experiments as their research methodology, ethical questions are especially apparent. This is also the case for interviews and surveys. All these research methodologies involve human beings in very direct and immediate ways. Of utmost concern in any of these methodologies should be that the subjects of study are treated with dignity and respect, that their interests are protected and that the potential costs of any study to the participant are weighed carefully against the potential benefits of the research to furthering knowledge and understanding. While Institutional Review Board (IRB) reviews are not a guarantee for an ethical research process, they, no doubt, are nevertheless a crucial necessity as they do formalize certain standard criteria that must be met to ensure the protection of the interests of those being studied.

The IRB review looks at several standard criteria in evaluating the appropriateness and ethics of a research study. Participants must give what is called **informed consent** to be included in an experiment, survey, interview, *etc.* Informed consent means that the subject is told that they are part of a research study; they are told what the general topic and purpose of the study is; they are told how the results of the study will be used (for example, published in a scholarly article that includes their name or maintains their anonymity); they are informed of the potential risks or costs to them as well as the benefits of participation in the study; and, finally, they are told they may discontinue their participation in the study at any time.

The goal of Institutional Review Boards to protect the rights of those being studied is particularly important in cases where the subject of study is considered technically unable to give consent (for example, when minors are studied (as was the case in the school voucher study discussed in Chapter 9) or when, for example, the mentally disabled are studied). True consent also depends on a

participant not being provided inducements to participate that are so great that they are unwilling or really unable to say 'no' to participation. For example, large monetary incentives for participating in an experiment or a survey are typically not allowed by Institutional Review Boards as they present so much coercive power to participate that the individual may not think clearly about the costs and benefits. One of the authors of this chapter (Elizabeth), for example, had a friend in college who agreed to participate in a medical experiment that subjected him to a treatable, but very disagreeable, bacterial disease because he needed the $500 they offered for his participation in order to pay his rent that month. The incentive very likely made him participate in a study he normally would have found too risky and costly.

Similarly, issues regarding full consent come to the forefront when the researcher holds some kind of power over the participant. For example, when a professor is conducting experiments and using some of her college students as participants, students may feel obligated to participate because they do not want to lose the favor of a teacher who may be grading them. Or, in the case of the school voucher study, for example, schools (both public and private) may feel unable to resist participation when they are strongly encouraged to participate by a state legislature that controls some valuable purse strings. Careful consideration must be given to provide fair but not overly coercive inducements or pressure to participate.

The costs or risks to a participant in a research study must be carefully assessed by the scholar conducting the study. Sometimes the subjects of the study themselves might not recognize the potential risk they face. For example, peasants in a Chinese village may unwittingly expose themselves to government backlash by sharing information with a Western researcher. Or, participants may be unwittingly confronted by survey or interview questions they did not expect to be asked and subsequently these questions cause them psychological distress.

Respecting the dignity of participants being studied means also that, whenever possible, participants should be given the opportunity to learn more about the study and the final findings. In the case of experiments, it is commonplace for participants to be debriefed at the end of the experiment. Usually in an experiment and often even with survey research and interviews, the researcher does not want to tell the participants at the beginning the full purpose of the

research or the hypotheses going in as they do not want that information to affect the participants' responses. For example, the field experiment in Chapter 11 was designed to investigate how Facebook "friends" can influence voting behavior. The participants in the study were not given the precise purpose of the study (finding out if Facebook friend posts can make individuals more likely to vote) because that information could have influenced the behavior of the participants. The researcher would be unsure if voting behavior changed because they knew what the researcher was trying to study or because of the posts. The consent form, therefore, included a more general, but still truthful, description of the study's purpose ("to understand the link between social networking sites and politics"). Similarly, in her interviews with survivors of the Argentinean military junta, Thalhammer in Chapter 5 did not want her interviewees to know that she was interested specifically in resistance because, as she said, she did not want them to tailor their answers to support (or disprove) her hypothesis.

Debriefing the participants after a study has been conducted on what the full purpose of the study was or sharing the published work with them respects their contribution to and sacrifice for the project, and indicates to the participants the valued role they played in furthering knowledge. At the same time, it also allows them to benefit from the knowledge gained. In the case of the school voucher study, for example, schools, state governments and non-profit organizations can use the results to help them consider the most effective educational policies to implement. Similarly, participants in a survey or in in-depth interviews may gain greater self-knowledge or understanding of how their opinions are similar to or different from others.

Occasionally, sharing the results of a study with participants can be difficult. For example, a researcher might present critical or unflattering information about a subject. In most cases, however, if a participant has given informed consent, has been allowed to maintain their anonymity of their own choosing, has not been unfairly coerced to participate, and the research has been conducted in a professional and objective manner, participants feel respected in the process and recognize the value of the knowledge gained by the study. Fair and respectful (even if critical) treatment of participants is important as well because scholars are part of a community of researchers. If one of us fails to treat our participants ethically and with respect, we burn bridges for other scholars. As a result,

scholars will then find it more difficult to get cooperation in their future projects and, thus, to continue to contribute to the scholarly body of knowledge.

One recent study by two political scientists underscores the importance of ethical methods to the ability to contribute to the scholarly body of knowledge as well as to the credibility of our research findings. Transparency in the methods one uses is important so that other scholars can **replicate** your findings, giving greater confidence that they are generalizable, and build on that knowledge. Increasingly, scholars are being called on to be transparent by making not only their methods but also their data accessible to other scholars for further analysis and confirmation. In 2014, a Ph.D. candidate, Michael LaCour, and a well-known professor at Columbia University, Donald Green, published the results of a field experiment that found that gay political canvassers could persuade people to change their positions on the issue of gay marriage. This research received significant attention in the field and in the media because it ran counter to past research which found that people were very unlikely to be persuaded to change political attitudes.[5] Two graduate students were so excited about these findings they began carefully examining the methodology and then the data used by LaCour and Green.[6] They discovered in the process that a significant portion of the data had likely been fabricated. LaCour, the main author of the original study, denied fabricating the data but could not prove his assertion because he said he had destroyed the original data – thus, violating the norms of transparency and the possibility for replicability. In addition, it was discovered that LaCour had falsely claimed a number of prestigious sources of funding in order to increase the perceived legitimacy of the study. Because of these ethical violations, the journal *Science* in concurrence with Professor Green (but not LaCour) retracted the article.[7] The retraction was covered by many media outlets and the unethical study did damage to the reputation not only of these scholars but also to the field of political science.

We want to conclude this section with an emphasis on objectivity as a normative requirement placed upon the researcher. Maintaining objectivity is a main imperative in doing research. It is easy to see how objectivity, as a principle, applies to all research methodologies, whether it involves human beings in direct and immediate ways or not.

From the outset, of course, objective collecting and handling of the data is

essential. It is especially important that scholars do not "cherry pick" the data or the participants in a study to support a particular hypothesis. For example, in [Chapter 10](#) on content analysis of media coverage of members of Congress, because of time and resource constraints, Danielle Vinson could not include all coverage of all members of Congress. However, to assure her sample was not biased, she carefully created a **random sampling** procedure, whereby no one type of Congressperson was more likely to be selected for her analysis than any other. Random sampling is similarly important in survey research. [Chapter 7](#) on statistical analysis discusses the Latino National Political Survey. It illustrates well how valuable such a data set was as all prior studies had been done on only small, non-random groups of immigrants and, thus, the results were not generalizable regarding why some immigrants choose to go through the naturalization process while others do not. Scholars of case studies, such as those discussed by Eyadat in [Chapter 3](#), must be particularly careful because of the small samples they often use to make sure their selection of cases is based on objective criteria of interest not just because they will help support some preconceived hunch or opinion, or the researcher's hypothesis or theory.

Objectivity also requires the researcher to maintain a certain level of scholarly detachment from the subject of study. As Brent Steele cautions us in his chapter on critical and interpretive research, "who we are as researchers, regardless of how 'objective' we may wish to be, impacts how we relate to subjects (whether those are films, or narratives, or the people we interview in our ethnographic methods), including the framing of the research questions used to explore the topic at hand." For example, especially in intense or personal interview situations or when doing field research that requires prolonged interaction, scholars may begin to identify with their subjects or develop a natural human connection. This kind of connection can be valuable as it may foster deeper insights into the subjects of study. However, it may also become more difficult for the researcher to evaluate her findings fairly. Kate Kaup gave us an example about field research in China: "If researching how political campaigns are conducted in China's countryside, for example, the researcher may come to know one of the candidates well and find it difficult to write about his or her failures, misconceived notions, or even corrupt practices." She urges: "Striking a balance between empathy and detachment is critical for presenting objective truths and

advancing knowledge."

Scholarly detachment is also important in unobtrusive research such as content analysis. Danielle Vinson, our author of the content analysis chapter, confided to us: "When analyzing the more subjective aspects of coverage, such as the tone or orientation of coverage, we must be careful to take what the author or speaker says at face value rather than reading our own views into the writing, and we must resist the temptation to impugn the motivations of the reporter because we disagree with what is written." When engaging in content analysis, or for that matter any other type of analysis, it is indeed important that we, as scholars, consider all possible explanations and avoid introducing our own biases – a danger and temptation that is always present.

Issues with One's Research Findings

We already hinted at the ethical dilemmas when publishing any findings from one's research in the case of Thalhammer's interview research on human rights activism in [Chapter 5](). The results may indeed carry serious implications far beyond the researcher's initial goals. This is also illustrated well by the disturbing results of a very interesting study. The researcher Keith Payne was interested in people's prejudices and their effects on perceptions.[8] In his experiment he primed the participants by showing them either a Black face or a White face. Then he exposed the experimental participants to very quick randomly sequenced images of a gun or a tool – two objects that have some level of resemblance especially when one can see them only for a split second. Every time the participants saw (or believed they had seen) the gun, they were to push a button. The finding was that participants misidentified tools as guns more often when primed with a Black face than with a White face.

One of us (Akan) first encountered this study in a graduate school when he took a seminar in Social Psychology. The reaction of many classmates was interesting and telling and it was expressed in the context of a very controversial court case that Keith Payne introduces his research article with. In 1999, four New York policemen shot and killed Amadou Diallo, an unarmed Black

immigrant from West Africa, in a hail of more than 40 bullets! Yet, the policemen were acquitted on the grounds that although they made a mistake, their actions were justified at the time. The shooting was judged to be justified because at the moment that policemen ordered Diallo to stop, the victim moved, producing an object that the policemen mistook for a weapon. Later on this object turned out to be a wallet. The police defendants contended that in this ambiguous situation, they acted on the information available, sincerely believing that they were in danger.

Keith Payne's study lends support to this conclusion. Moreover, many students in the seminar concluded immediately that the results deliver a scientific basis (or, more problematic, a *justification*) for the existence of prejudice and racism. They had serious normative reservations about this study and the publishing of its results. Yet, the research was done objectively and it followed all ethical principles of scientific research. It is then in the researcher's discretion to evaluate the pros and cons, the benefits and the possible negative implications of publishing the findings. Although Akan was initially shocked and had serious reservations about Keith Payne's publishing of these results, he is now in favor of such types of studies as they shed light on important fundamentals of human interactions. We need knowledge, however uncomfortable and disturbing it may be. Knowledge is a first and indispensable step toward effective measures against racism or whatever we may face.

Political science researchers must also consider whether their research methodology could extend beyond their original goals of gaining knowledge to even more significant outcomes such as changing the political world. Most recently, two political scientists conducted a field experiment to examine whether giving voters information about the ideology of judicial candidates could influence vote choice. Using the natural experiment of an actual election, mailers were sent to a large portion of Montana voters (15%) describing the judicial candidates as situated somewhere between Mitt Romney and Barack Obama on an ideological scale. Given the size of the mailing, many people worried that the study itself may have influenced the outcome of what was supposed to be a nonpartisan election. The study was also criticized for ethical violations for using the official seal of the State of Montana on the mailer without permission of the state, a violation that was done in an attempt to make the experimental materials

appear official to the participants in the field experiment.[9]

Finally, we want to emphasize what should be obvious, namely that researchers do not introduce their own biased interpretations when presenting the results. Researchers are human beings and human beings have their own opinions and biases. While the goal of political science research is providing objective knowledge, it would be disingenuous to suggest that a researcher's own interests and sense of what matters does not affect their decision to pursue a certain research question. However, objectivity and the standards of scientific study require that the conclusions we draw once the facts have been gathered are not influenced by our own feelings or biases. As they were reflecting on their survey research from Chapter 8, Lyman Kellstedt and James Guth remarked to us:

> [One] potential ethical problem was in the interpretation of results: religion involves important value issues on which people differ strongly. As researchers we are not immune to such considerations and might be tempted to interpret some data in a way that is critical of religious traditions or groups other than our own.

It is probably the case that most of us start our research program with an expectation of what we will find. Not finding one's expectations confirmed by the research results may be disappointing at first. However, discovering that our preconceived notions are incorrect can be one of the most rewarding and interesting parts of the whole process. It is a process of learning – a process that (hopefully) never ends, and the researcher should be as open to it as he or she expects the readers of the research to be.

Final thoughts

Throughout this book, you have heard firsthand from scholars using a variety of the methodological tools available in political science. These tools include the comparative case study, field research and interviews, statistical and survey research, document and content analysis, experimental research, and interpretive

research. We have presented these methods in neatly packaged chapters. This may have created the impression that they are distinct and mutually exclusive from one another. The astute reader probably has noticed by now, however, that these methods are not exactly distinct but, instead, overlap in fundamental ways. For example, we have a standalone chapter on statistical research, but you likely noticed that most methodologies, including experiments, document analysis, content analysis, sometimes in-depth interviews and, always survey research, rely on statistical analysis. Similarly, though presented in separate chapters, field research is not entirely distinct from interviews as it typically involves open-ended, in-depth interviews.

Additionally, researchers know that often the very best way to answer a research question is to rely on multiple methodologies. Using multiple methodologies, scholars can gain in their investigation both the depth offered by such approaches as open-ended interviews and the breadth offered, for example, by large-scale surveys. Sometimes one scholar or one research team uses multiple methodologies to address their research question. If the results obtained through one methodology are confirmed by the results that have been obtained through another methodology, then we gain increasing confidence in the outcomes. The results validate each other, in other words.

Finally, research is a cumulative and sometimes messy process. Though science has the goal of finding truth, often our conclusions and findings are modified over time as we learn more and continue in the quest for truth. Sometimes we pursue false leads, as the authors of Chapter 8 did in their survey research on religion and politics. Sometimes we get frustrated in our pursuits, as Kaup was by the Chinese government's limits on her access to certain groups in her field research discussed in Chapter 4. Often, we find that research can be time consuming, as it was in the in-depth interviews in Argentina presented in Chapter 5. Occasionally, research can require considerable methodological training and skills, as is apparent in the statistical research presented in Chapter 7. But, if you have discovered anything from reading these chapters in the authors' own words, to a true scholar doing political science research is always, at the very least, an interesting and, at the very most, an ultimately rewarding pursuit.

Exercises and Discussion Questions:

1. Should researchers be allowed to fund their research with monies from private organizations?
2. Can you think of research topics that are "off-limits"? Why are they not (to be) pursued?
3. Why do you think that issues of ethics are typically only discussed in the last chapter of most research methodology books?
4. Which research methodologies, do you think, bring the least amount of ethical challenges with them? Why?
5. Do you think scholars can truly be objective? If not, what are the implications?

Recommended Resources:

Office for Human Research Protections (www.hhs.gov/ohrp/index.html): Learn more about Institutional Review Board guidelines and policies as set forth by the U.S. Department of Health and Human Services Office for Human Research Protection.

NIH Office of Extramural Research (www.phrp.nihtraining.com/users/login.php): Take an online training course in the ethical use of human subjects created by the National Institute of Health (this training is often required to be completed by Institutional Review Boards before approval for research is granted).

American Political Science Association (www.apsanet.org/content_9350.cfm): Download the American Political Science Association's Guide to Professional Ethics, Rights and Freedoms.

Center for Media and Social Impact (http://cmsimpact.org/code/code-best-practices-fair-use-scholarly-research-communication/). See also the Code of Best Practices in Fair Use for Scholarly Research in Communication.

Notes

1 Humphreys 1970.

2 Henle and Hubbell 1938: 230.

3 Tunyaz 2018.

4 Zimbardo 2007.

5 Carey 2015.

6 Aschwanden and Koerth-Baker 2016.

7 McNutt 2015.

8 Payne 2001: 81.

9 Willis 2014.

References

Aschwanden, Christie, and Maggie Koerth-Baker. 2016. "How Two Grad Students Uncovered an Apparent Fraud – And a Way to Change Opinions on Transgender Rights." https://fivethirtyeight.com/features/how-two-grad-students-uncovered-michael-lacour-fraud-and-a-way-to-change-opinions-on-transgender-rights/

Carey, Benedict. 2015. "Study Using Gay Canvassers Erred in Methods, Not Results, Author Says." https://www.nytimes.com/2015/05/30/science/michael-lacour-gay-marriage-science-study-retraction.html/

Henle, Mary, and Marian Hubbell. 1938. "Egocentricity in Adult Conversation." *Journal of Social Psychology* 9(2): 227–234.

Humphreys, Laud. 1970. *The Tearoom Trade: Impersonal Sex in Public Places.* Chicago: Aldine Publishing Company.

McNutt, Marcia. 2015. "Editorial Retraction." http://science.sciencemag.org/content/348/6239/1100.2

Payne, Keith. 2001. "Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon." *Journal of Personality and*

Social Psychology 81(2): 181–192.

Tunyaz, Tohti. "China/Xinjiang Uighur Autonomous Region." https://pen.org/advocacy-case/tohti-tunyaz/ (accessed 3/13/2018).

Willis, Derek. 2014 (October 28). "Professors' Research Project Stirs Political Outrage in Montana." https://www.nytimes.com/2014/10/29/upshot/professors-research-project-stirs-political-outrage-in-montana.html?_r=0

Zimbardo, Philip. 1971. "The Power and Pathology of Imprisonment." *Congressional Record* (Serial No. 15, 10–25). Hearings before Subcommittee No. 3, of the Committee on the Judiciary, House of Representatives, Ninety-Second Congress, *First Session on Corrections, Part II, Prisons, Prison Reform and Prisoner's Rights: California.* Washington, DC: U.S. Government Printing Office.

Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil.* New York: Random House.

GLOSSARY

A

Absolute law: (see **law**).

The American Political Science Association (APSA): the largest professional association for political scientists. Formed in 1903, this association holds annual professional conferences where scholars share their research. The APSA also publishes the premier journal in the field, the *American Political Science Review.*

American politics: a subfield in the discipline of political science devoted to understanding American politics and institutions.

Analysis of variance (ANOVA): a statistical technique that compares the means of two or more groups on some variable to see if they differ significantly.

Antecedent variable: occurs in time prior to an independent variable and may act as a catalyst or a precondition for the independent variable to take effect.

Attrition: the process through which subjects drop out of a study over time. This happens mostly in panel, experimental and field studies. For whatever reason, subjects may no longer want to participate, they may move away, *etc.*

B

Behavioral revolution: the name given to the shift in research emphasis among political science scholars beginning around the 1950s from a focus on describing political institutions to a more scientific, systematic, data analysis-driven study of politics and political behavior.

Bivariate analysis: statistical analyses that examine the relationship between

two variables, usually one independent variable and one dependent variable.

C

Case study: an in-depth investigation of a case, where the case may refer to an individual, group, phenomenon or event.

Causal mechanism: a statement that explains *why* there is a relationship between two or more variables; it is the explanation for *why* something happens.

Causality: in order to establish causality three criteria must be met: first, the assumed independent variable and the dependent variable must correlate; second, the independent variable must precede the dependent variable in time; third, the relationship between the independent variable and the dependent variable must not be **spurious**.

Closed-ended questions: questions that force respondents to pick among a set of response options.

Comparative method: a systematic method for discovering empirical relationships among variables with the goal of establishing general empirical laws.

Comparative politics: a subfield in the discipline of political science devoted to explaining politics by comparing political institutions, most often nation-states, using either the method of similarity or the method of difference.

Concepts: the words that describe the phenomena under investigation.

Concepts must be **conceptualized**, which simply means that they must be defined.

Conceptualization: the verbal definition of a concept.

Confidence interval: the range of likely values of a variable in the population.

Confidence level: a statistic representing how confident one is that the sample value represents the population value of the variable.

Content analysis: a method of research used to analyze communications such as speeches, news stories, etc., typically quantitatively.

Content analysis categories: exhaustive and mutually exclusive measures used

in content analysis to evaluate the presence or absence of variables of interest.

Content code: the numeric value assigned to a variable that is of interest in a content analysis.

Control group: members in the randomly selected group that are assigned to not receive the independent variable expected to change the measured dependent variable. The effect of the independent variable is measured by comparing individuals in the treatment group to individuals in the control group.

Control variables: variables that are held constant in quantitative analysis to isolate the relationship between the variables of interest.

Controlled comparison design: a method used by researchers to compare members of groups who share similar demographic characteristics in order to help alleviate the problem of selection bias in a sample.

Convenience samples: samples of respondents used because they are readily available.

Correlation: a measure of how two variables are related to each other. A correlation can be positive (as one variable increases in value the other does as well) or negative (as one variable increases the other decreases), but correlation does not prove a causal relationship between the variables.

Cross-sectional survey: when a sample is surveyed at one point in time only.

Crosstabulations: a statistic which shows the extent to which two variables occur together.

Cumulative: a characteristic of scientific knowledge which evolves in reference to previous scientific findings and either improves and complements them or shows them to be erroneous.

D

Data: systematically collected and objective observations about the phenomenon being studied; it can come in two forms: quantitative or qualitative (numerical or textual).

Dependent variable: the presumed effect in a causal relationship.

Descriptive statistics: basic statistics, including frequencies and measures of

central tendency (mean, median, mode), which describe the data.

Dichotomous measure: a measure that can be one of two categories.

Dummy variable: a variable that equals 1 if it is in a particular category and 0 if it is not.

E

Effect size: estimated effect of a one-unit change in an independent variable on the value of the dependent variable.

Elite interview: the process of interviewing decision-makers in the area under investigation.

Empirical evidence: data-supported proof that a proposition is true.

Empiricism: a way of knowing the world and occurrences in it through sense experiences and observation, that is through the process of data-gathering.

Episodic record: secondary data that captures a particular period of time; examples include papers from a Supreme Court justice, a presidential diary, *etc.*

Epistemic correlation: a theoretically grounded, but empirically (still) unsupported relationship between a concept and the variable measuring that concept.

Experiment: a method used by researchers to study causal relationships where the researcher in a controlled setting manipulates an independent variable to test its effect on a dependent variable.

Experimenter effect: changes to responses by individuals in a study caused by perceived cues from the researcher, such as the desirable response to a survey question or to perceived expectations.

External validity: the degree to which methods or findings are generalizable outside of the study.

Extraneous variable: a factor that could have created an "accidental" or "non-causal" relationship between the independent and dependent variables.

F

Factorial design: a study that simultaneously examines the effect of two or more independent variables.

Failure-to-treat-problem: when individuals assigned to the treatment do not successfully receive the treatment, such as when they cannot be contacted to receive a message.

Field experiment: a naturally occurring experiment where the researcher has no control over who is in the treatment group and who is in the control group.

Fieldwork: the researcher is collecting data through observation or interviewing in a natural setting.

Film analysis: a micropolitical and interpretive method that examines the political themes and lessons within films and how those both reflect, but also influence, political discourses and debates. Film analysis pays particular attention not only to what is said, but also what is excluded and/or demonstrated through film's visual and aesthetic representations.

First differences: in a logit regression, a statistic representing the likely increase or decrease in the dependent variable if each independent variable is raised from its minimum to maximum value while holding the other variables constant.

Formal model: the mathematical and abstract representation of the phenomenon under investigation.

Frequency: how often something occurs – for example, how often a category in a content code occurs in a particular sample.

G

Generalization: the process by which the findings about one event or occurrence are made applicable to all events or occurrences of the same class.

H

Hawthorne effect: when participants in a study act differently because they know they are the subjects of a study.

Hypothesis: an educated guess about the direction of a relationship between an independent variable and a dependent variable.

I

Independent variable: the presumed cause in a causal relationship.

Index: an additive combination of ordinal variables, all measured at the same level and identically coded, producing a composite measure of a complex political phenomenon.

Inferential statistics: the process of drawing generalizations from the analysis of a sample to the entire population.

Informed consent: when a participant agrees to be in a study knowing fully the risks and benefits associated with such participation.

Institutional Review Boards: panels created to review the ethics of research being proposed and ensure that the participants in the study are protected.

Intent-to-treat effects: a comparison of differences in the dependent variable of interest between members of the treatment and control groups that does not (or cannot) account for the failure-to-treat problem. This is often used when actual contact cannot be measured.

Inter-coder reliability: the extent to which multiple readers using the same content analysis procedures agree in placing the content into coding categories. It tests the validity of various measures of content.

Internal validity: the extent to which the causal claims found in the study represent true cause-and-effect relationships.

International relations: a subfield in the discipline of political science devoted to investigating the relationships among various international actors, including nation-states and other political organizations.

Interpretive methods: methods that translate data into meanings. Interpretive

methods include narrative, discourse and textual analysis, and ethnographic research. Such methods assume that the "data" of politics does not present itself unproblematically and thus must be interpreted to "make sense."

Interval level variables: variables which have categories which are equidistant from each other but no true zero (e.g., temperature).

Intervening variable: an intermediate factor between the independent and dependent variables in a causal chain; it is caused by the independent variable and it causes the dependent variable.

L

Law: a correlation between two variables. We distinguish between **absolute law** and **probabilistic law**. An absolute law states: "Whenever we observe x, we will observe y" or "If x, then always y." A probabilistic law states: "Whenever we observe x, we will observe y, with the probability z."

Literature review: a section of a research paper that reviews the existing scholarly findings regarding one's research question with the goal of establishing why pursuing the research question is important.

Logistic regression: a statistical analysis of experimental data that includes a dichotomous (only two values) dependent variable. The generated estimates for each independent variable refer to the log odds change generated in the two possible values of the dependent variable.

Logit regressions: a statistical method that shows the impact of several independent variables on a dependent variable that is dichotomous or has only two possible values.

Longitudinal analysis: analysis done on data that is collected over a significant period of time; can be one of three types – **trend analysis**, **intervention analysis**, or a **panel study**.

M

Margin of error: the range around which a sample statistic may fall (above and below).

Matching: a process used in a controlled comparison design where members of the groups being analyzed are chosen because they are similar on certain relevant characteristics; this allows the researcher to isolate the effect of the independent variable on the dependent variable.

Measurement error: the extent to which an instrument fails to measure accurately what it is intended to measure; for example, when a thermometer misreads the air temperature or when a survey question is interpreted by the interviewees to be asking something different from what the researcher intended.

Meta-analysis: a review of combined data from a number of individual experiments generating an average treatment effect estimate.

Method of agreement/similarity: a research design by which cases are selected that all have the same outcome; at the same time, the cases must be dissimilar in regard to all but one presumed independent variable.

Method of difference: a research design by which cases are selected that all have different outcomes; at the same time, the cases must be similar in regard to all but one presumed independent variable.

Methodological pluralism: using multiple complementary methods to study a particular question. It allows researchers to gain a more complete understanding of the issue.

Micropolitical analysis/research: research conducted at the level of everyday subjects, including churches, schools, cafes, town hall meetings, sporting events, as well as through films or other forms of popular culture. Micropolitical research assumes politics found at broader levels of societies and states is both reflected and reproduced as well as modified through the social interactions of humans.

Models: statistical equations that describe how a dependent variable is related to one or more independent variables.

Multi-item index: a single-scale variable created by combining several different survey items believed to measure different components of the same concept with the goal of reducing measurement error.

Multivariate analysis: data analysis in which the effects of multiple independent

variables on the dependent variable are examined.

Multivariate logistic regression models: statistical analyses of experimental data that includes a dichotomous (only two values) dependent variable and multiple independent variables. The generated estimates refer to the log odds of the two possible values of the dependent variable.

Multivariate regression analysis: a form of statistical analysis using multiple different independent variables in order to see the relationship between each of those variables and the dependent variable while holding the other variables in the model constant.

N

N: the symbol used to represent the number of people or entities in a sample.

Narrative extrapolation: a discourse, used or reinforced by anyone from elites, to social scientists, to everyday citizens, that helps to specifically tell a causal story, functioning to "teach" individuals the logic regarding particular policies.

Nominal level variables: also called categorical; variables which have unordered categories (e.g., gender, religion).

Normative/Normatively/Normative question: suppositions regarding how one believes things ought to be.

Normative question: a question about how things ought to be in an imagined world. An example of a normative question would be: how should a country be governed?

Null hypothesis: the hypothesis that there is no relationship between the independent and dependent variables.

O

Open-ended questions: questions that allow respondents to give their own answer as opposed to being forced to choose among pre-created response

options.

Operationalization: turning a concept into an indicator or measure that can be analyzed; the process that lends a concept measurability.

Ordered logistic regression: a statistical procedure used when the dependent variable is ordinal; it estimates the impact the independent variables have on the *likelihood* of observations being in the next highest category of the dependent variable.

Ordinal measures: measures that are ordered from low to high in value (for example, from lower class to upper class), but where the increments between the values are not necessarily equivalent or definable.

P

P-value (or statistical significance): a statistic that shows how likely a value from a sample reflects the true population value.

Panel study: a type of longitudinal analysis where the researcher measures variables of interest from the same sample over a period of time.

Participant observation: the observation of any group under interest while being a 'member' of the group. It allows for the development of intimate familiarity with the group.

Peer-review: the process by which a scholar's work is reviewed and evaluated by academic experts on the subject before possible publication.

Pilot study: a small preliminary survey conducted to evaluate new or alternative questions or improve other aspects of survey design before conducting a full-scale research project.

Political theory: a subfield in political science. It concerns itself with philosophers and is normatively driven.

Population: the total membership of any defined units of analysis. It can refer to people (like in a presidential approval poll), objects or events.

Positive question: a question about phenomena in the existing world. The opposite of a positive question is a **normative question**.

Positivist/positive statement: an approach to understanding that relies on facts,

data and the objective world.

Postbehavioralism: a reaction to the behavioralism of the 1950s–1970s in which methodological concerns seemed to outweigh the substance of study. Postbehavioralism alerted scholars to the need for policy relevance in their studies.

Post-test only, **control group experimental design**: an experimental design where measurements of the dependent variable are taken only after the intervention in order to avoid testing/**instrumentation effects**.

Pre-test/post-test, **control group experimental design**: the classic experimental design where participants are randomly assigned to either an experimental group or a control group; the value of a dependent variable of interest is measured in both groups before and after an intervention or manipulation is introduced on the experimental group to test the effect of the intervention on the dependent.

Primary source analysis: research that examines data that are directly collected or observed by the researcher, such as when a researcher interviews elected officials or observes a political protest.

Probabilistic law: (see **law**).

Program evaluation: an assessment of the outcomes produced by a public policy; often there is a focus on whether or not the program or policy is effective in meeting its stated goals.

Proposition: an expression of a judgment or a declaration about the relationship of at least two concepts; it must be either true or false.

Published data: data that is generally publicly available and has been collected by a government entity, private organization or other researchers; examples include Census data, Congressional roll call votes and National Election Studies survey data.

Q

Qualitative analysis: non-numerical evaluation of data.

Qualitative methods or research: research that is non-numerical; in-depth

research or data collection often limited to a few cases.

Quantitative analysis: numerical evaluation of data.

Quantitative methods or research: statistical or numerical research that works by assigning numbers to the terms that are being studied.

Quasi-experimental: designs used by researchers who are interested in estimating a causal effect of a variable on an outcome, but are unable to use random assignment to treatment and control conditions for practical or ethical reasons. Generally, these designs have lower internal validity than experiments.

Quota sampling: a type of stratified sampling where researchers define categories of interest (race, gender, etc.) and then select, typically in a non-random way, a sample of respondents in each category proportional to the population of interest.

R

Random assignment: the process in an experiment where individuals are assigned to either the experimental group or control group; assignment is completely by chance.

Random sampling/selection: a sampling procedure in which each element (e.g., individual) has an equal probability of being selected for the study sample. This results in a sample that is representative of the population.

Randomized control trial: a research design (also known as a field experiment) in which subjects are randomly assigned to treatment and control conditions. Unlike lab experiments, randomized control trials take place in a natural, real-world setting.

Rapport: to gain a certain amount of trust from an interviewee so that s/he feels comfortable being engaged in the interview and answering the interviewer's questions.

Ratio level variables: variables that possess full mathematical properties (age, income).

Reactivity: when a subject in a study behaves differently because they are

reacting to being the subject of study.

Refereed journal: a scholarly journal containing research papers which have been reviewed anonymously by other scholars before publication to ensure quality research has been done.

Reliability: the extent to which a measurement measures something the same way, time after time.

Replication/Replicable: the duplication of a scientific study with the purpose of exploring the possibility that the original results are invalid because of particular research conditions and/or procedures.

Research design: the plan for answering one's research question using a particular research method.

Research question: a statement that identifies the phenomenon we want to study; it is generally motivated by curiosity about something that we consider important but has not been asked, addressed and answered yet – at least not satisfactorily.

Response set: when questions are worded in such a way that respondents tend to answer "agree" to every question.

R-squared statistic: a measure of the goodness of fit of a statistical model; how well the independent variables account for or explain the variation in the dependent variable.

Running record: secondary data that is collected at regular intervals over a period of time.

S

Sample: a subset of cases drawn from a population.

Sampling error: a statistic that represents the difference between the sample value and the population value for a certain variable, calculated based on the size of the sample.

Scientific knowledge: when objective and systematic observation of data that can be verified by others is used to explain or understand something.

Secondary analysis: uses existing data for purposes other than those for which

the data was originally collected, usually to address a new research question or provide an alternative perspective on the original study.

Secondary data: data collected by someone other than the researcher.

Secondary data analysis: analysis of data that was not originally collected by the researcher; this data might include autobiographies, newspaper photos, legislative records, *etc.*

Secondary source analysis: research that examines data that is collected by indirect, secondhand or unobtrusive methods; for example, a researcher uses secondary source analysis when she examines the text of political speeches or analyses Congressional roll call votes.

Selection bias: the bias in data that occurs when the sample is non-representative of the population under investigation.

Selective deposit: a term used in document analysis to describe the bias caused by only certain information being collected (such as only a well-funded group being able to afford keeping detailed records).

Selective survival: a term used in document analysis to describe the bias caused by only certain documents being preserved over time.

Semi-structured interview: the interviewer has an 'interview guide' of topics or issues to be covered; however, s/he is free to change the wording and order of questions and to follow topical trajectories that, if appropriate, may stray from the guide.

Snowball technique: asking those interviewed to suggest names of potential interview subjects with particular characteristics, viewpoints or opinions.

Spurious relationship: the assumed independent variable and the dependent variable are in fact not causally related, if there is, in fact, a third (hidden) **extraneous variable**.

Standard error: a statistic reflecting the difference between the sample value and the population value.

Statistical analysis: the use of mathematics to explain data.

Statistical inference: generalizing the results from a sample to the larger population of interest.

Statistical significance (or p-value): a statistic that shows how likely a value from a sample reflects the true population value.

Structured interview: consists of a list of specific, pre-set questions; the

interviewer strictly adheres to the list and the order in which the questions appear and does not introduce any spontaneous questions or comments into the interview process.

Survey research: a method used to study public opinion where a random sample of a population is asked questions which are then compiled and analyzed statistically.

T

Theory: a statement or a series of statements that explain(s) *why* there is an association between two or more concepts (variables). It is related, but distinct, to a law which, more simply, states *that* an invariant or probable association exits.

Traditional political science: practiced in the pre-1950s, political science was largely descriptive (as opposed to analytical), and specific and narrowly focused as opposed to nomothetic or focused on general laws.

Treatment group: members in the randomly selected group that are assigned to receive the independent variable expected to change the measured dependent variable. The effect of the independent variable is measured by comparing individuals in the treatment group to individuals in the control group.

U

Unit of analysis: the object or the entity under study; for example, in a content analysis, the unit of analysis might be an entire news story or individual paragraphs. It may also refer to individuals or larger aggregates, such as political, ethnic, religious or any other grouping or countries.

Unstandardized coefficients: coefficients in a regression equation expressing in the units of the variable itself the effect of a one unit change in the independent variable on the value of the dependent variable, controlling for any other independent variables in the model.

Unstructured interview: questions are not pre-set; the interviewer elicits information (data) from the interviewee in a free-flowing conversation.

V

Validity: the extent to which a variable is measuring what is intended, typically evaluated by looking at whether that measure is related in expected ways to other variables.

Verification: the process of demonstrating the validity or truthfulness of a hypothesis or a theory by means of collecting empirical data.

INDEX

Note: Page numbers in **bold** type refer to figures
Page numbers in *italic* type refer to tables
Page numbers followed by 'n' refer to notes