Newsletter

APSA - CP

Volume 14, Issue 1
Winter 2003

The Organized Section in Comparative Politics of the American Political Science Association

Contact
Decio Hall, Box "D"
University of Notre Dame
Notre Dame, Indiana 46556
Tel. 574-631-5681
Electronic correspondence preferred
http://www.nd.edu/~apsacp

Letter from the President

Comparative Politics and the Real World

Evelyne Huber
University of North Carolina
ehuber@unc.edu

Since September 11, there have been many calls for social science to be more relevant, specifically to contribute to a prevention of terrorism by furthering our understanding of its causes. These calls have stimulated a discussion about the larger question of what social science, or more narrowly political science and comparative politics, owes society or can contribute to informed political choices (e.g. Diamond, PS Online Special, March 2002; www.apsanet.org/PS/post911/diamond.cfm). I would like to contribute to this discussion by suggesting that there are at least two questions to consider when we talk about the policy relevance of comparative politics. The first question is what scholars in comparative politics can really do, and the second is how likely political systems are to absorb this knowledge and provide the channels for its application in practice.

To answer the first question, we have to start with taking stock of how we understand the political world and of the areas in which we have acquired substantial knowledge. By the same token, we should also reflect upon the limits of the capacity of social scientists to engage in political engineering. There are plenty of examples of seemingly reasonable political initiatives gone awry. For the purposes of this discussion, it is useful to distinguish between analyses of complex, macro-level phenomena, of the functioning and effects of political institutions, and of the nature and effects of specific policies or combinations of policies.

A very significant part of comparative politics is concerned with macro-level processes and outcomes such as transitions among regime forms, revolutions, models of capitalism, or models of development. All of these are multidimensional phenomena and the product of the interaction of many variables. For instance, to explain revolutions, analysts at a minimum look to changes in the social structure, in the nature of the state apparatus, in political alliances, and in a country's relationship to the international economy and political system. We have very good explanations of the constellations of these variables that have produced revolutions, but scant hard and fast specifications of necessary or sufficient conditions and their interactions. Much the same could be said about successful models of development, or the formation and survival of democratic regimes in the developing world. Thus, we clearly cannot design any blueprints for successful development models, or for the construction and consolidation of democracy.[1] The best we can do is to identify the constituent policies of these models and to estimate the probabilities that they could be replicated in different historical contexts.

(Continued on page 2)

Letter from the Incoming Editors

Michael Coppedge
University of Notre Dame
coppedge.1@nd.edu

Anthony M. Messina
University of Notre Dame
messina.3@nd.edu

It is with considerable enthusiasm, humility and more than a little apprehension that we assume the responsibilities for editing, producing and distributing the semi-annual Newsletter of the Comparative Politics section of APSA. Our overarching goals are to maintain and expand upon the excellence of the Newsletter in representing within it the full spectrum of methodological approaches to the study of comparative politics and its many discrete research and pedagogical traditions and agendas. To achieve these goals, we require and, thus, welcome practical suggestions from our readers.

As in previous issues, we will include in this and future Newsletters the *Letter from the President*, *News and Notes*, *Symposium*, and *Controversies*. We also offer one revised section and two new ones.

The format of the revised book review section, under the stewardship of Kathleen Collins, will vary from one issue to the next. One model will have several scholars, representing different research or methodological traditions, reviewing an especially prominent book. A second model, as represented by the essay in this issue by Lisa Anderson, will be a single-authored review of several books on a common subject or theme. Finally, as in the past, we will occasionally solicit single-authored reviews of one book.

Our first new section informs readers about the availability of new empirical data sets. This section will describe the contents of the data sets and provide links to relevant source pages.

Another new section of the Newsletter raises questions of pedagogy. This section will explore various approaches to teaching graduate students field and other research techniques. It will also occasionally offer practical strategies on how best to teach comparative politics to undergraduates.

In both continuing and departing from previous practice we wish to acknowledge the invaluable contributions of the previous editor, Daniel Treisman, his Newsletter staff and their predecessors in making the Newsletter a vehicle of information dissemination and stimulating intellectual discussion. With the assistance of our Editorial Board, we will try our best to extend their tradition of excellence.

(Continued from page 1)

Another major and growing part of comparative politics is concerned with understanding the political consequences of institutional designs. These analyses are somewhat less ambitious in terms of the scale and complexity of the outcomes they attempt to explain than analyses of regime changes, development models, etc., and therefore lend themselves potentially more easily to the formulation of policy prescriptions. Yet, within the study of institutions, there are also differences between areas where there is a stronger consensus and knowledge is more precise, and other areas where there are more contentious scholarly debates and knowledge is thinner. For instance, we know quite a bit about the institutional underpinnings of clean elections, rules on financing of parties and campaigns that help level the playing field, and about the consequences of electoral rules for party discipline. We know much less about the institutional characteristics that make and keep bureaucracies efficient and free from corruption.

Various international and U.S. actors have sent delegations to observe elections in many countries, and these delegations are guided by norms about institutional procedures considered prerequisites for clean elections. The verdict of such external observers is generally given much credence in the international community, as it is relatively easy to identify violations of institutional norms. For instance, leaving aside the campaign for the moment and focusing on the election only, the secrecy of the ballot, fraud-proof identification procedures, the presence of poll watchers, and opening hours long enough to accommodate all those seeking to vote are considered key norms in the conduct of clean elections. However, if we want to make the step from judging adherence to institutional norms to formulating prescriptions for institution building, we need to ask whether these are necessary conditions. Let us just consider identification procedures. They may be necessary in contexts where election fraud is widespread or recent, but when I go to vote at the Chapel Hill Public Library I am amazed every time that no identification is required. I simply tell them my name and address and they check me off on the voter roll and hand me a ballot. Presumably, if someone else showed up claiming to be me, after I had already voted and been checked off, they would charge that person with attempted voter fraud. Do we conclude, then, that this is a luxury that an old democracy with highly institutionalized and well-functioning election procedures (after Florida in 2000 ?!) can afford, or do we conclude that a high probability of being detected if attempting fraud would be an effective deterrent elsewhere, and thus complicated (and often very costly) identification procedures are not necessary, as long as you have good voter rolls and a functioning system of recording who voted? This would certainly be an important question to answer for a would-be engineer of electoral institutions in the real world, where scarcity of resources is an important factor.

How to construct a bureaucracy that is efficient and not corrupt has preoccupied rulers and social scientists for a very long time. Two generally accepted institutional prerequisites are merit-based recruitment and promotion. However, it is less clear whether these are also sufficient conditions, or what else is required. For instance, much of the literature used to judge the South Korean bureaucracy, which is by all accounts highly meritocratic, to be a model of honesty, but more critical examinations in the wake of the East Asian financial crisis have thrown considerable doubt on this assessment and thus on the assumption that meritocratic recruitment and promotion are sufficient conditions. Another approach is exemplified by the Independent Commissions Against Corruption in Singapore and Hong Kong, which are credited with having reduced corruption very significantly in these two contexts where corruption was historically a big problem. Can we conclude that similar commissions with high levels of independence from the authorities of the day, a permanent status, and broad mandates extending to prevention would be successful in other contexts, by themselves or only in combination with merit-based recruitment and promotion? There are also suggestions that strong parties and alternation in power are conducive to low corruption; do they constitute an alternative to Independent Commissions Against Corruption? Are both alternatives equally effective?

What are the implications of this discussion of our understanding of the consequences of institutions for our willingness and capacity to make policy recommendations? In general, we know what has worked in the past, and we try to explain why it worked in a certain way and to generalize about the relationships between specific institutional characteristics and outcomes of interest. The reach of the generalizations we make depends on our approach. Rational choice approaches tend to assume that institutions structure incentives such that actors with the same motivations will act in the same way under the same institutional configurations. Accordingly, those analysts who assume thick rationality (the pursuit of power or money) claim that their generalizations about the consequences of institutions are valid regardless of context. This then yields the strongest policy recommendations for institutional engineering. Multivariate statistical analysis, in contrast, attempts to control for context and to identify the amount of variation explained by the institutional variables of interest and the probability that the findings are statistically significant. Accordingly, the generalizations are more modest, claiming a high probability of effects of a specific magnitude for institutional engineering. Moreover, the use of controls and interaction terms implies that if we put the same institutional elements together again in a different context, their functioning and the results will be the same only if there is no new relevant contextual factor that our original theory and analysis overlooked. Comparative historical analyses of the functioning and effects of institutions tend to result in the same probabilistic and context-sensitive generalizations. On this basis, only cautious policy recommendations for institutional engineering can be made.

Arguably, the part of comparative politics that concerns itself with the causes and consequences of specific policies is the one that lends itself most readily to policy recommendations. Here we have a considerable amount of both positive and negative knowledge, that is, we have accumulated experience with policies that have worked well and with others that have worked less well. For instance, we know that social security systems, including pensions, sick pay, and unemployment benefits, based on a combination of a universalistic flat rate with universalistic earnings-related benefits result in the lowest rates of poverty and in the most equal distributions of income. We also know that generous child and family allowances (paid to the mother) combined with the availability of subsidized public day care are effective means to keep single mothers, a particularly vulnerable group, out of poverty. In health-care policy, we know that systems based most heavily on private delivery and financing are most expensive and most inegalitarian, both in developed and in developing countries. And among policies designed to support gender equity in economic terms we know that separate taxation, active enforcement of anti-discrimination provisions, availability of subsidized public day care, and entitlement to social security benefits through part-time work are all effective, particularly if used in combination.

We also know something about policies that are characteristic of successful development models, such as investment in human capital through public health systems and quality primary education. We know further that land reforms and investment to improve the productivity of small and medium scale family farming contribute to lessening of poverty and inequality and to improvement in human capital in developing countries. And we know that such policies have to be financed in a fiscally sustainable manner, which in turn requires the implementation of an effective system of taxation.

Our knowledge extends further to the conditions under which various policies tend to be adopted. For instance, we know that in advanced industrial democracies the political power distribution, particularly partisan incumbency, which in turn is related to the strength of organized labor and the history of religious cleavages, is crucial for the adoption of generous welfare state policies and for their redistributive impact. Both social democratic and Christian democratic parties, if in power for prolonged periods, build generous welfare states that are effective in keeping people out of poverty, but social democratic incumbency results in more redistribution and less social inequality. Or, the combination of strong women's movements and social democratic incumbency is most effective in advancing gender equity in economic respects. This knowledge is essential for our understanding of how the world of politics and policies works, and it brings us back to the question of the kinds of policy recommendations that are likely to be implemented in different contexts. But before addressing this question, let me try to summarize what we can say about the capacities and limits of comparative politics in generating policy recommendations.

As noted above, our capacities are strongest in the formulation of recommendations for specific policies to achieve specific goals via the allocation of resources (e.g. to universalistic social safety nets to reduce poverty and inequality) or the provision of services (e.g. public health care or day care to facilitate quality of access and quality), or via regulations of social interaction (e.g. non-discrimination and affirmative action provisions to reduce discrimination against women and minorities). They are weaker when it comes to recommendations for the building of political institutions, though there are differences here between different kinds of institutions. They are weakest, or non-existent, when it comes to recommendations for bringing about macro social change, though for some of these macro social outcomes, such as successful development models, we can identify constituent policy elements, such as investment in human capital, and formulate corresponding recommendations. Arguably, our most valuable contributions are to lay open empirical regularities as well as exceptions to those regularities, and to expose policy makers to the variety of experiences with policies and institutions. On the basis of these experiences, we can then narrow the probability for policy makers that our recommendations will achieve the desired results.

Clearly, the most dangerous undertaking, and one scholars had best stay away from, is to invent a new institution, based on certain assumptions about human motivations and behavior. We simply cannot predict human behavior, most fundamentally whether people will play by the institutional rules we design, or whether they will attempt to circumvent those rules in legal or illegal ways. Let me illustrate this with a little example from field research years ago - and I am sure many readers could offer similar stories. The Peruvian military government under President Velasco (1968-75) decided to attempt a class conciliation project in the urban industrial sector, in the hope that this would increase productivity and let unions wither away. For this purpose, they decreed a law that made all employees of an enterprise members of an Industrial Community, which was to receive a share of before tax net profits to reinvest in the enterprise, along with representation on the board of directors of initially two members, to grow with the share of ownership. In addition, individual employees were to receive a share of profits in cash. The expectation was that this would lead to more reinvestment, labor peace, and a loss of influence of unions. With very few exceptions, the exact opposite happened. Entrepreneurs were outraged by the legislation, and rather than reinvesting in their enterprises themselves to slow down the share of collective ownership by the employees, they decided to circumvent the law by minimizing declared profits. Workers realized that they were being cheated and turned to unions for help. Rather than diminishing, the number of unions in the industrial sector more than doubled between 1968 and 1975.

But even if we stay on firmer ground, confining ourselves to drawing generalizations from observed empirical regularities and making statements about the probability that certain combinations of actions, or policies, will result in desired outcomes, our chances of seeing these outcomes effectively realized may be slim. The reasons for this are obvious: The power distribution in society and the consequent orientation of policy makers has to be favorable for the adoption of given policy recommendations; in other words, the policy makers have to be willing to pay the costs associated with the policies that will bring about the results that they profess to pursue. Let us just consider one example: Policy makers in the United States as well as Latin American countries profess a commitment to a significant reduction of poverty. From the experience of Western European countries, we know how this can be done. Social scientists can generate (and have generated) estimates of the percentage of GDP that would need to be devoted to transfer payments and social services, specifically health and education, and from there it is an easy step to assess the extra revenue that the government would need to generate. We also know that government revenue in the United States is way below the average for advanced industrial democracies, and that Latin American countries have a lower tax burden than other countries at comparable levels of development. But here, of course, comes the problem. Politicians would need to be willing to confront those who could pay more taxes, the upper middle and upper income groups, and those are precisely the groups in the United States and in Latin American countries that also have the most political influence, given the lack of strong parties representing lower middle, working and lower class interests.

Where does this leave us with regard to the question of what comparative politics owes to society? Certainly with the obligation to continue to accumulate knowledge about the shape and consequences of policies, institutions, and macro social change, regardless of current interests among policy makers. I disagree, for instance, with the advice offered at a conference on economic and social policies in Latin America that European experiences with basic pension systems before or immediately after WWII, when some of these countries were at similar levels of development as some Latin American countries are now, are irrelevant to study because "nobody in policy making circles talks about this any more." The reason why nobody in policy-making circles talks about this has to do with the power of the proponents of neoliberal doctrine. One of the privileges we have as social scientists in an advanced industrial democracy is to buck the trend and speak truth to power. In order to do this effectively, we owe it to ourselves and to society to pursue our research into questions about differences in the functioning and consequences of policies and institutions, and about macro social change, with customary high scientific standards. Clearly, one can choose questions with more or less relevance for the real world of politics, and my personal preference would be for more, but we should make those decisions on the basis of our own interests and values, not on the basis of what is currently of interest in the circles of power.

Notes

[1] The fact that some people are making lots of money and enjoying great prestige as consultants, peddling one-size-fits-all blueprints for successful development, should not detract our attention from this truism.

Studies in Democratization

The Center for the Study of Democracy at Northeastern University is pleased to announce the online publication of its new ejournal, Studies in Democratization.

Studies in Democratization is a biannual ejournal dedicated to publishing meaningful contributions to the fields of Democratization Studies and Comparative Politics. Submissions are welcomed from both established and new scholars.

The first edition includes four articles. They are: "Civil Society and African Democratization: The Flip Side of The Coin" by Shadrack Wanjala Nasong'o; "Civil-Military Relations in the Republic of China: a Conceptual Approach to Evaluating the Stability of Taiwan's Civil-Military Dialogue" by Peter Richardson; "Oil and the Role of the Major Powers in the Caspian and Central Asia" by Itir Toksöz; and "West-East Civil Society Partnership: Partnering for Growth or Preparing for Conflict?" by Paul Beran II.

Visit www.csd.neu.edu for more information and to view the ejournal.

Call for Nominations

The commitees for the Luebbert Book and Article Awards, the Sage Paper Award, and the Data Set Award solicit nominations for each of these prizes. Please send your nomination to the members of the relevant committee (see below).

Committee Members for the Comparative Politics Section

Nomination Committee

T.J. Pempel, University of California, Berkeley (chair), pempel@socrates.berkley.edu
David Cameron, Yale University, davidrcameron@yale.edu
Valerie Bunce, Cornell Univerity, vjb2@cornell.edu
John Carey, Washington University, Saint Louis, jmcarey@artsci.wustl.edu
Jeffrey Herbst, Princeton University, herbst@princeton.edu

Luebbert Book Award Committee

Margaret Levi, Cornell University (chair), mlevi@u.washington.edu
Gretchen Casper, Penn State University, gcasper@psu.edu
Richard Snyder, University of Illinois, rsnyder@uiuc.edu
Please inquire from Professor Levi where to send copies of books.

Luebbert Article Award

Gary Marks, UNC-Chapel Hill(chair), gwmarks@unc.edu
Gerardo Munck, University of Illinois, g-munck@uiuc.edu
Deborah Yashar, Princeton University, dyashar@princeton.edu

Sage Paper Award

Michael Wallerstein, Northwestern (chair), m-wallerstein@northwestern.edu
Sofia Perez, Boston University, sofiap@bu.edu
Matthew Shugart, University of California, San Diego, mshugart@ucsd.edu

Data Set Award

Gary Freeman, University of Texas, Austin (Chair), gfreeman@jmail.la.utexas.edu
Michael Alvarez, De Paul University, malvarez@depaul.edu
Ron Inglehart, University of Michigan, rfi@umich.edu

French Politics

We are pleased to announce the forthcoming launch of a new journal in political science, *French Politics*, to be published by Palgrave. The first issue will appear in March 2003. The journal aims to focus on all aspects of French politics from the full range of theoretical and methodological perspectives common to comparative political science, and submissions will be particularly welcomed that incorporate the French case in both comparative and cross-national analyses. The journal will also feature an innovative data section, that will be accessible both in print form and electronic form to subscribers. The multi-national executive editorial board includes John Gaffney (University of Aston), Vincent Hoffman-Martinot (University of Bordeaux/CNRS), John Huber (Columbia University), Amy Mazur (Washington State University), Anand Menon (University of Birmingham), and Andy Smith (University of Bordeaux). The data editor is Michael Lewis-Beck (University of Iowa) and the book review editor is Catherine Fieschi (University of Nottingham). For further details, please contact the editors, Andrew Appleton (Washington State University) and Robert Elgie (Dublin City University) at appleton@wsu.edu or elgie@dcu.ie. Guidelines to authors, as well as subscription information, may be found on the journal website at: www.palgrave-journals.com/fp.

A New APSA Organized Section for Qualitative Methods

Colin Elman, Arizona State University
David Collier, UC Berkeley
Henry E. Brady, UC Berkeley

A diverse group of APSA members has launched an initiative to form a new Organized Section for Qualitative Methods within the American Political Science Association.

If approved by the APSA, the new section will sponsor research and training in the several branches of methodology associated with the qualitative tradition. It will likewise focus on the relationship between these diverse methods and other areas of methodology, including quantitative methods.

Cooperation with the APSA Comparative Politics Section and with the *APSA-CP* Newsletter will be important to the new section, given their interest in methodological issues and the major contribution of comparativists to recent debates on methodology. The new section will complement the activities of the APSA Political Methodology Section by emphasizing the qualitative side of methodology, as well as seeking to develop productive avenues of cooperation with the existing methods section.

In addition to sponsoring APSA panels and short courses, as well as a newsletter, the Qualitative Methods Section will co-administer the two-week intensive Training Institute held annually by the Consortium for Qualitative Research Methods (CQRM), located at Arizona State University. The Training Institute scheduled for January 2003 will accommodate 60 students who will be taught by 20 faculty, with both students and faculty drawn from universities across the United States. For further information on CQRM and the Training Institute, see http://www.asu.edu/clas/polisci/cqrm/ .

The petition proposing the formation of the Qualitative Methods Section, endorsed by over 1,000 APSA members, has been submitted to the American Political Science Association for approval. The petition is available at http://www.asu.edu/clas/polisci/cqrm/qualsect.htm. For more information, please contact Colin Elman colin.elman@asu.edu or David Collier dcollier@socrates.berkeley.edu. For information about the newsletter, contact John Gerring jgerring@ias.edu.

Bridging the Quantitative-Qualitative Divide

Introduction

In 1994, King, Keohane, and Verba's *Designing Social Inquiry* exhorted scholars to hold qualitative and quantitative research to a common set of methodological standards. In the subsequent eight years we have witnessed high-profile debates (in this Newsletter and the *American Political Science Review*) about what the common standards should be and which approach is best able to satisfy them. Although this debate has been stimulating and useful, we believe it is more constructive to focus on ways to bridge the qualitative-quantitative divide. In this spirit, we identified several research strategies that would exploit the complementary strengths of qualitative and quantitative research. We then combed the literature, searching for scholars who have skillfully executed these strategies. In some areas, such as dealing with measurement error in qualitative research, we decided that little progress has been made so far. But in several other areas, scholars have developed innovative ways of bridging the divide. In this symposium, five scholars who are actively bridging the divide describe their strategies, reflect on their experiences, and recognize others who are pursuing similar strategies.

Leading off, Stathis Kalyvas explains how he used both quantitative and qualitative evidence to best advantage in explaining the location and timing of political violence during the Greek civil war. Next, Amy Mazur and Dorothy McBride Stetson describe the challenges of creating quantitative indicators of complex concepts, using the example of their collaborative effort to measure characteristics of quite diverse women's movements in Western Europe. Third, Michael Ross explains why the hypothesis that oil hinders democracy, developed most notably by Middle East area specialists, could be tested properly only with a larger, cross-regional sample. Fourth, Evan Lieberman shows how "nesting" case studies in a large-N quantitative analysis can improve theory development, case selection, inferences, and interpretation. Finally, Robert Franzese notes that many hypotheses in comparative politics involve some causes conditioning the effects of others. He argues that such hypotheses can and should be tested with interactions.

The Logic of Violence in Civil War

Stathis Kalyvas
University of Chicago
kalyvas@uchicago.edu

I am currently studying the dynamics of violence in the context of civil war (my manuscript in progress is *The Logic of Violence in Civil War*). A central task is explaining spatial variation in rates of civilian victimization (in civil wars, civilian victimization tends to exceed battle deaths, usually by far). In designing this project, I faced three challenges.

The first was conceptual: the very concepts and categories used to describe violence are derived from everyday discourse and journalism and are, thus, potentially misleading. The term genocide (and its various uses) is a case in point, while the very concept of "political violence" is laden with a host of implicit and problematic assumptions that make both analysis and testing very difficult. The same is true for more complex, yet widely accepted concepts, such as ethnic conflict or ethnic violence.

The second problem was factual. Our understanding of the phenomenon of civil war violence is based on stylized facts (e.g. neighbors attacking each other, or people killed in large groups) that may or may not be true. The reason we don't know is the lack of basic systematic research. Likewise we tend to generalize backwards (e.g. ethnic identities as they crystallized at the end of the conflict are assumed to exist at its start) and impute motivations to actors from their alleged identities (e.g. violence among members of different ethnic groups is called ethnic violence without further investigation). Such stylized assumptions underlie much research, ranging from case studies and macro-historical analysis to formal and econometric models, thus contaminating their findings.

The third problem was related to data. Most available indicators of "political violence" tend to be unreliable and inconsistent across nations and over time. The reasons are well known. Data are difficult to collect in times of war, poor countries (where civil wars tend to erupt) have poor or no record-keeping institutions - and the war usually destroys whatever capacity administrative existed in the first place. Fatalities data are usually politically salient, hence the parties to the conflict have strong incentives to bias and manipulate them while restricting access to them. The bias is generally systematic, with perpetrators reducing the numbers and representatives of those who have been victimized exaggerating them. Often, fatality figures tend to become canonical and are therefore hard to revise. In the case of Bosnia the canonical figure is of 200,000 fatalities, a number that emerged in 1993 from the Bosnian Information Ministry, a party in the conflict. Independent estimates are closer to 60,000 fatalities (keep in mind that Bosnia is one of the most intensively covered civil wars, both by journalists and by scholars). Furthermore, there is a systematic tendency to underreport violence from remote rural areas (I refer to this as "urban bias") as well as forms of violence even more difficult to measure, particularly sexual violence. Indicators for identities (including ethnic ones) also tend to be extremely problematic, for a variety of well-known reasons.

Because of the data problem I made a decision which eventually proved beneficial as it allowed me to address also the conceptual and factual challenges. I opted for a subnational research design, with villages as the units of analysis rather than states. This research design has two disadvantages. First, it is limited by the range of its empirical basis, and second, the relevant causal processes cannot be generalized across levels: the variables that explain the spatial variation of violence across villages would not explain the same variation across nations. Nevertheless, these disadvantages are offset by substantial gains: the design allows for the rigorous test of hypotheses and the generation of robust facts and concepts. The focus on microdynamics makes sense once violence is distinguished analytically from conflict; and hypotheses about the variation of violence within states can be useful in providing insights about the same variation across states. When complemented with non-systematic comparative evidence, rigorous subnational empirical analysis offers considerable advantages, particularly in important yet underdeveloped and data-poor fields of investigation.

Collecting data on civil war fatalities proved to be a highly labor-intensive process. Altogether, I spent one year and a half conducting archival research and interviews in a Greek rural area where a civil war raged in 1943-1944. I studied two counties that included 64 villages with a total population (in 1940) of 45,140 inhabitants. The civil war resulted in fatalities that totaled close to 2 percent of the population. By collecting the data myself, I was able to reconstruct the events that took place in this region in great detail, as well as code every single instance of violent homicide in the region as an event in a given space and time, surrounded by its political, social, cultural, and institutional context, and placed within a sequence of events. All observations came with a "story," including the identity of the perpetrator and the victim, the time and location of the homicide, the way it was carried out, the links between perpetrators and victims, the links between this homicide and both anterior and subsequent instances of violence (or the lack thereof), the justifications (if any) that were given (or that can be inferred) for the homicide, and, when possible, the decision-making process that preceded it. Most observations were corroborated from multiple oral and written sources. In addition, I was able to collect less systematic data on other forms of violence, such as arrests, beatings, rapes, etc. This data collection procedure produced important insights about the nature of violence. For example, one concerned the identity of perpetrators: even though most acts of violence were perpetrated by "specialized" groups of armed men, many more ordinary civilians were involved through the process of denunciation. This insight turned out to match a widespread aspect of violence in many civil wars, one which is absent from existing macro accounts of violence, ethnic conflict, revolutionary mobilization and collective action, and civil war. Last, an important aspect of the research design is that it avoids the selection bias which, until recently at least, had plagued the study of violence.

Based on intuitions from fieldwork, a wide comparative reading, and insights from various theoretical literatures (including among others military history, criminology, and the anthropology of feuding), I developed a theory and specified a model from which I derived predictions about the spatial variation of violence. The model's central insight is that selective violence is largely a function of the degree of territorial control exercised by the competing political actors. Empirically, this means that certain levels of control should be associated with certain levels of violence. Hence, the most straightforward use of the data was to see whether they conformed to the predictions of the model. In addition, I used a number of sociological, economic, political, and demographic controls to make sure that the observed variation was not due to other variables.

The use of both qualitative and quantitative data and tools offered several advantages. I will discuss five significant ones.

First, qualitative information collected from interviews, archives, and local histories were indispensable in constructing the indicator for the key causal variable, territorial control. Because territorial control takes on fine gradations, items such the frequency of visits by various armed groups or the presence of clandestine organizations had to be compiled through various sources, both oral and written.

Second, to increase the robustness of my findings and introduce additional controls, I was able to design additional "out-of-sample" tests, using qualitative data from other Greek regions. For instance, because my indicators for territorial control tended to overlap somewhat with ecological variables (e.g. mountainous terrain), I used observations in the vast literature of local history and civil war memoirs to make additional inferences. To give an example, a prediction of the model is that high levels of rebel control should result in low violence; because high levels of rebel control tend to coincide with areas of high altitude, I could use observations in the literature about villages located in mountains and not experiencing violence as indirect additional evidence. At the same time, because this relative overlap could be an indication of bias (e.g. low violence in mountains caused by traditions of mountain solidarity), I was able to locate villages where rebel control and altitude did not overlap, so as to discriminate between the two.

Third, interviews offered the possibility of live process-tracing, so to speak. Several of my early hunches and intuitions met their end when confronted with ethnographic or archival evidence. Alternatively, I felt confident that these intuitions were correct when my informants offered information and interpretations that corresponded or were consistent with the hypothesized mechanisms. For example, one key mechanism posited by my model is the transmission of information from individuals to organizations by means of denunciation. However, denunciations are not directly "observable" either because they were not recorded or because the relevant sources were lost. However, villages are contexts where information about such processes is generally of good quality and I was, therefore, able to collect reliable information on denunciations. Moreover, I was able to locate individuals who personally handled denunciations and were willing to talk to me. In this way, I was able to verify that the mechanism I hypothesized was indeed present. On a more informal level, I was even able to use some of my informants as sounding boards for ideas, intuitions, and hypotheses: after all, these were people who had gone through the very process I was studying. At the same time, some of the facts offered by my informants proved to be wrong when confronted with archival information or evidence provided by other informants. What is more, my informants' own interpretations about various issues could be checked against quantitative data. For example, when asked why their villages had been relatively peaceful when neighboring ones experienced considerable violence, many of my informants pointed to their village's tradition of moderation and solidarity, a tradition that was often confirmed by many other people, including the inhabitants of surrounding villages. Because I was able to collect (mostly judicial) data on prewar village conflict, I was able to test this claim and found it wanting: prewar conflict did not predict violence or its absence during the civil war. In other words, the effect of a tradition of moderation and solidarity was most likely endogenous to the non violent outcome: because these villages escaped violence (through other mechanisms than moderation and solidarity), they were able to construct a postwar tradition of moderation and solidarity which was then used to explain in an ex post facto way the absence of violence. This example underscores the power of a simultaneously qualitative and quantitative investigation using multiple sources of evidence.

Fourth, and related to my previous point, the qualitative dimension of the investigation offered the possibility of "testing" implications of the hypotheses that were difficult to quantify. Such implications included the motives of denunciations (whether they were malicious or not), the risk aversion and age of denouncers in particular areas, and the relevance of personal characteristics of local leaders. Although it was impossible to collect quantitative evidence about these variables, it was nevertheless possible to collect reliable qualitative evidence. Such evidence was not systematic, but it was usually possible to ascertain whether it was credible and reliable or not through interviews and ethnographic investigation. Obtaining qualitative evidence about the implications of the hypotheses increased my confidence in the quantitative results.

Fifth, the combination of qualitative and quantitative evidence made possible a very productive analysis of outliers. Though my model did a generally good job in predicting the spatial variation in violence, it also failed in a number of cases. There were both errors of over-prediction and under-prediction of violence. Armed with this information, I returned to the field and investigated the outlying villages, collecting additional information and conducting additional interviews in order to understand what "went wrong." Obviously, this investigation would have been impossible in the absence of prior quantitative testing. This research proved highly useful. For instance, the determinants of non-violence in villages where the model predicted violence are different from the determinants of non-violence in villages that were correctly predicted by the model. These outliers have a significance that is not just analytical but carries potential policy implications: by understanding how villages remained peaceful even under the kinds of incentives and constraints that should have produced violence in the first place, I was able to reach a deeper understanding of the "non-parametric" causes of non-violence, which are precisely those that may turn out to be more amenable to actual manipulation through outside intervention. Moreover, this research carried theoretical significance. Since the model assumed instrumental behavior, the outliers allowed me to isolate non-instrumental processes, such as emotions and norms. Unlike studies that seek to identify emotions and norms inductively and are, thus, potentially vulnerable to the charge of post hoc induction, my analysis locates them in a more robust way, via the failures of the model.

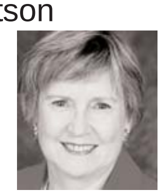
The social scientific study of "political violence" is on the verge of a dramatic leap precisely because of the imaginative combination of qualitative and quantitative tools and data, along with more rigorous analysis. This is most clearly exemplified by recent and forthcoming work on ethnic riots, most notably by Ashutosh Varshney (*Ethnic Conflict and Civic Life: Hindus and Muslims in India*. New Haven: Yale University Press, 2002) and Steven Wilkinson (*The Electoral Incentives for Ethnic Violence: Hindu-Muslim Riots in India*. Forthcoming, Cambridge University Press).

* Thanks to Elisabeth Wood for helpful comments.

Quantifying Complex Concepts: The Case of the *Women's Movement* in the RNGS Project

Amy G. Mazur
Washington State University
mazur@mail.wsu.edu

Dorothy McBride Stetson
Florida Atlantic University
stetsond@fau.edu



The Research Network on Gender, Politics, and the State (RNGS) has recently embarked on the task of converting its data-rich qualitative case studies into a valid and reliable quantitative data set for dissemination to scholars interested in women's movements and state feminism. As such, it can be a useful case study of bridging the qualitative-quantitative divide in Comparative Politics research. It was a rude awakening when we discovered that, despite careful attention to developing a research design usable by the 45 researchers in the network we were not prepared to operationalize the central concept of the study--women's movements. Here is the RNGS plan:[1]

1. Developing the state feminism model, a theoretical framework for understanding the impact of women's movements and women's policy agencies on state responses;

2. Defining key concepts and variables and applying common indicators for the collection of data;

3. Selecting individual policy debates as the units of analysis in five issue areas (abortion, prostitution, job training, political representation, and a 'hot' issue of national priority);

4. Selecting and analyzing a sample of three debates in each country for each issue area and one debate for the hot issue;

5. Comparing qualitative country-based analyses of policy debates in five books, one for each issue area.

With the qualitative studies nearly completed, the network has embarked on developing a code book for quantitative measurement. We are pursuing this bridging strategy for several reasons. The size and expertise of RNGS makes it possible to amass a large enough set of data-rich, theoretically structured cases to test research hypotheses statistically. Submitting the qualitative observations to the rigor of the coding process will provide greater confidence in the findings with respect to the model of state feminism. In addition, putting the findings in quantitative form will give scholars, practitioners, and activists easy access to detailed information on policy making processes over a 30-year period, across five policy areas, in 14 countries and the European Union.

The RNGS qualitative research plan provided an initial set of nominal definitions of concepts and lists of indicators to guide RNGS researchers in investigating and reporting on their cases. To complete the codebook, it has been necessary to review each of these definitions with an eye to creating quantitative measures of the original qualitative concepts treated in the case analyses. This article covers our work on the central and most difficult concept of the study-- the women's movement.[2] We describe the key challenges in the context of current scholarship and describe the strategies that we have adopted to overcome them.

Challenges in Measuring Women's Movements

In the qualitative phase of the project, the network allowed individual researchers to describe and classify the movement in their respective countries. Their studies produced a wide range of entities with little cross-case consistency that can be broadly classified as national networks, organizations, individuals and groups inside institutions, and grass-roots groups. Such variation made it difficult to apply rigorous quantitative standards in measuring the features of movements that figure prominently in the model of state feminism (see n. 2). In preparing to bridge qualitative and quantitative approaches, it was imperative to develop a more focused understanding and more precise definition of women's movements. Our ultimate goal, therefore, was to generate a working research definition that would be comparative, empirical, valid, and reliable.[3] To elaborate:

1. The definition must be applicable cross-nationally with minimal conceptual stretching (Sartori 1970; Collier & Mahon 1993).

2. The nominal and operational definitions must enable researchers to distinguish, through observation, phenomena that represent the concept from those that do not, but also allow for variation cross-nationally and longitudinally (Collier 1998).

3. To establish validity, both the nominal definitions and the operational definitions must yield observations that coincide with expert understanding of the concept (Kaplan 1964).

4. To establish reliability, the operationalization and measurement of phenomena that represent the concept must be applied in similar ways by a variety of researchers.

We learned early in the process that there is little consensus over a definition of women's movements (Beckwith 2000; Molyneux 1998). We searched the literature to develop a useful definition or, failing that, to compose a common working research definition. We were unable to find any specific definition, but we were able to construct a composite definition:[4] Women's movements are *groups and associations using disruptive and conventional tactics to change women's position in society*. This definition was consistent with RNGS qualitative studies, met the criteria of comparability, and, since we study post-industrial democracies, could be applied without conceptual stretching. The problem with this definition was that it lacked validity, that is, it did not encompass all the ways that experts use the concept. Although many scholars study organizations, they admit that, in doing so, they are not really covering the entire women's movement (Ferree and Hess 1996). Others find the movement, not in organizations but inside formal institutions (Katzenstein 1998; Santoro & McGuire 1997).

Another, smaller, set of scholars suggested a shift away from actors toward ideas in defining a movement. According to Mansbridge (1996: 27):

> Women's movement...is a set of changing contested aspirations and understandings that provide conscious goals, cognitive backing and emotional support for each individual's evolving feminist identity.

Along the same lines, Rochon (1998) sees a movement existing in the relation between a critical community of intellectuals who develop and articulate new identities and values and the actors who bring these ideas to social and political life. According to Mansbridge and Rochon, then, the movement can be discerned in thoughts, voices and spaces between and among persons, organizations and communities. Defining the movement in terms of ideas rather than organizations moves us closer to meeting the criterion of "face" (Gurr 1972) or "conceptual" (Collier 1998) validity by reflecting expert assessments that movements are more than the organizations that advocate for them. At the same time, it moves us away from comparability and reliability across cases, since it is virtually impossible to study ideas apart from the individuals and organizations that advance them. Jenson (1996) provides the final piece of the puzzle, enabling us to move toward the measurement of movement characteristics and activities. In her work on the French feminist movement, she distinguishes between politics as representation of the self to others via a collective identity from representation of interests and presentation of ideas to the state. In so doing, she separates the discourses of feminism from the actions of feminists.

RNGS Strategies: Focusing on Women's Movement Actors

Our response to the challenge of developing an operational definition of women's movements was, on the one hand, to take the ideas approach and define the concept as a set of discourses, beliefs, opinions, and identities developed by communities of thinkers about women. On the other hand, we also clarified the object of our research to be the *women's movement actors*, not the women's movement, and focused our definitions and measurements on the individuals and organizations that present and represent women's movement discourses, beliefs, opinions and identities in public life. To assess the advantages and disadvantages of this strategy, we describe here its contribution to the operationalization of two movement-related variables in the state feminism model that are central to the study: the impact of women's movements on policy content and the closeness of the movements to the political Left.

First, a note about our bridging procedures: we follow a similar process to operationalize each qualitative concept for quantitative analysis.[5] We begin by examining the definitions and indicators for each concept in the state feminism model and compare summary values given by researchers for each debate for cross-case reliability. Based on these comparisons, the second step is to propose measures that remedy problems of cross-case reliability. This frequently means that we unpack complex concepts into their separate components for measurement, producing several variables pertaining to each concept. After consultation with researchers, the next step is to develop detailed worksheets for recording information for each variable from the published case studies. For data not available in the chapters, we return to the researchers to provide missing information. Once completed, we each code the data separately from these worksheets and compare the results to assure cross-coder reliability.

Impact on Policy Content - In this concept, we are interested in whether policy outcomes reflect the ideas of women's movements, a key indicator of movement success (Gamson 1975). Before, in the absence of a valid definition of women's movement, we were unable to determine which goals to use as a standard for determining success. By focusing on women's movement actors, we can now distinguish two measures: the types of actors involved in each debate and their goals for state action in the debate, thus enabling us to measure which goals of which actors coincided with the policy outcome. We assign a 1/0 for the presence or absence of each of these: informal organizations, sectors of political parties or trade unions, organized interest groups, and individuals in institutions. The goals of the actors are coded by comparing each goal to elements of the policy outcome. Using this list of groups to represent the movement minimizes conceptual stretching by covering the full range of women's movement actors in postindustrial democracies. Researchers can distinguish women's movement actors from other policy actors as well. The measure is valid on its face given the nominal definition of women's movement actors, and it yields measures that can be tested for reliability. The initial inventory of these actors and movement goals will serve as the basis for coding many of the other characteristics of women's movement actors.

Closeness to the Left - This variable is one of the seven characteristics of the women's movements that might explain their differential impact on state policies. The nominal definition refers to ideological and organizational closeness to political parties and trade unions associated with the left. Left-wing parties are defined here in a programmatic sense, rather than in terms of any fixed position on a universal ideological continuum. That is, in countries where parties are aligned along a left/right spectrum, left parties are those that are more likely to promote agendas that emphasize greater political, social, and economic equality than parties of the right. We have also included trade unions when they are important left-wing actors.

Without a research definition of women's movements, however, researchers found it difficult to assess closeness. Initial classifications by researchers on this variable lacked cross-case reliability. The shift to a focus on actors allowed us to zero in on just which actors had attachments to the Left and how strong these bonds were. Closeness to the Left is now measured at three different levels. *Very Close*: Some women's movement actors formally ally or work with political parties and/or trade unions of the Left; ideas from the movement are taken up by left-wing parties in party platforms; and activists have internal power positions in the left-wing parties or unions. *Close*: actors formally ally or work with political parties and/or trade unions of the left but they do not have positions of power within them. Moreover, if the Left takes up the ideas offered by movement actors, they do so without bringing these ideas into the party platform. *Not Close*: movement actors and left parties and/or unions are remote from or hostile to one another. The quantitative codes for these levels are 2/1/0.

Assessments and Conclusions

Unlike the contested concept of democracy, which expert literatures have defined to death with little consensus (Collier and Levitsky 1997), the women's movement has been under-defined. The strategy developed by RNGS to overcome the under-definition of such a thorny concept has definite advantages. Defining the women's movement in terms of the actors that express a specific set of ideas and discourse satisfies our criteria for a good working research definition: avoiding conceptual stretching and achieving validity and reliability. As such, it supports the larger argument that concepts and measures constructed from small-N qualitative analyses are much closer to the micro reality of observed phenomena than aggregate numerical indicators that are often mere proxies for phenomena under study (Ragin 1994). That is, in the interest of selecting a single numerical indicator for highly complex concepts - a quantitative substitute for a qualitative concept - the very essence of the notion can be severely clouded or even lost. Notable examples include operationalizing democracy in terms of voting rates or women's political rights with women's presence in the legislature. This is not to mention larger problems with the validity of aggregate datasets that are used for quantitative cross-national analysis.

Distilling the goals and activities of concrete actors from qualitative descriptions and locating them in terms of the different organizational forms also permits accurate, yet cross-nationally portable analysis. The RNGS strategies for bridging the qualitative/quantitative phases of data collection also promise greater confidence in the measures' reliability. Conducting structured case analyses with a refined theoretical model and publishing them in the five policy issue books, re-examining the validity and reliability of qualitative measures in terms of those analyses, re-working measures with an eye for quantification, and then checking the new coding schemes are all integral parts of the RNGS bridging process. The usefulness of this dialogue may now be scrutinized by others who are searching for a protocol for converting qualitative studies into data sets for quantitative analysis.

Notes

[1] The project has involved extensive collaborations and consultations among network members. For more on the RNGS saga, project, and preliminary findings consult www.rngs.org; contact the authors, who are also the network conveners (mazur@mail.wsu.edu; stetsond@fau.edu); or see Mazur (2001); Stetson (2001); Stetson & Mazur (2000); Mazur & Parry (1998).
[2] Out of 22 variables in our original state feminism model, 13 involve some aspect of the women's movement. Seven variables are part of the larger cluster called women's movement characteristics: Stage, Closeness to the Left, Priority, Cohesion, Location, Feminist Activism, Countermovement, and Strength. The other four related variables measure the degree to which women's movements have been represented, both descriptively and substantively, in policy debates--whether the content of policy decisions that end individual policy debates coincide with movement goals and whether movement actors participate in policy debates. We also determine whether leaders have a record of involvement in women's movements.
[3] This report is part of a larger working paper, Stetson & Mazur 2002, which is available on request.
[4] See, for example Vargas & Wieringa (1998); Bull, Diamond & Marsh (2000); Basu (1995); Molyneux (1998); Beckwith (2000); Katzenstein (1987). See Stetson and Mazur (2002) for a complete review of the literature.
[5] We could not have completed this procedure without the key advice of James Caporaso, Laurel Weldon, Michael Mintrom, Jacqui True, Ashley Grosse, Janine Parry, Andrew Appleton, Diane Sainsbury, and Gene Rosa.

Bibliography available online on the *APSA-CP* website

Testing Inductively-Generated Hypotheses With Independent Data Sets

Michael Ross
UCLA
mlross@polisci.ucla.edu

Scholars in comparative politics often fail to conduct true out-of-sample tests. Many instead use a single data set both to generate hypotheses and "test" them. Of course, the resulting "tests" are not tests at all: hypotheses that are generated inductively from one data set can only be tested with a different data set. The problem afflicts both qualitative and quantitative studies.

There are many strategies for avoiding this problem. Qualitative researchers who use one case study to generate a hypothesis can conduct new case studies to test it. Alternatively, they can look for new data within their existing case - perhaps at another level of analysis - to test one of their hypotheses' observable implications.[1] Quantitative researchers can seek out data sets that cover new countries or new time periods; or if they are already employing global data sets, they can subdivide them to approximate out-of-sample estimations (one version of this latter approach - dubbed "the Stanford test" - is explained in Gary King and Langche Zeng, "Improving Forecasts of State Failure," *World Politics*, July 2001).

Of course, these strategies may be costly and are seldom employed. Too often scholars use an alternative strategy: they disguise their inductive hypotheses as deductive ones, and use the same case studies, or quantitative data sets, that they employed to produce the hypothesis to falsely "test" it. As a result, our field is strewn with hypotheses that have been dressed up as findings. As the saying goes, "it's not what you don't know that hurts you, it's what you know that ain't so." Much that we "know" in comparative politics probably ain't so because we fail to carry out valid out-of-sample tests.

One reason these disingenuous practices are common is that we place an absurdly high value on generating original hypotheses. Many of us want to want to be theoretical innovators; few are satisfied to test existing theories. Yet there are plenty of influential theories - or at least, hypotheses - waiting to be tested in the world of comparative politics, and testing them with independent data sets has many advantages.

This is what I did in my recent article, "Does Oil Hinder Democracy?" (*World Politics*, April 2001). The article takes a claim that is common among Middle East area specialists: that oil wealth inhibits the rise of democracy. It discusses the argument's origins, reviews the case-study literature, and then tests it with pooled time-series cross-sectional data for 113 states between 1971 and 1997. In short, it finds that Mid East specialists were right: a state's reliance on oil exports *does* tend to make it less democratic, both in the Middle East and elsewhere. It also finds that one of the theory's implications - that other types of mineral exports will have a similar effect - is valid.

Three causal mechanisms seem to account for this pattern. First, mineral-rich governments tend to use low taxes and patronage to dampen democratic pressures. Second, these governments spend exceptionally large sums on the military, which implies they are better able to repress dissent. Finally, mineral-based development appears to depress employment in a country's manufacturing and service sectors, which in turn hinders the development of democratic movements.

I was drawn to the arguments of the Middle East specialists (even though, by training, I am a Southeast Asianist) since I am interested in the problems of resource-rich states generally, and these scholars had thought carefully about issues that interested me. At the same time, I also learned to tread carefully. The first time I presented a draft of my paper at a conference I was bombarded by criticism for idly speculating about the influence of Islam on democracy.

Still, drawing on the hypotheses of area specialists had many advantages. Area specialists commonly harbor a wealth of generalizable insights about politics, insights that are often overlooked by those of us trained in formal modeling or statistical analysis. Many of us have learned these skills because we believe they will help our work become more rigorous, and help the discipline become more scientific. If we believe in the accumulation of knowledge, however, it is absurd to ignore the scholarship of others even if it has not been expressed in formal or statistical language.

Insights that are incubated by area studies specialists often have implications that reach far beyond the region's boundaries. Hypotheses about the oil-rich "rentier states" developed by Middle East specialists had important implications for oil-rich and mineral-rich countries in other regions; they even had implications for our understanding of democratization. Yet they were rarely discussed outside the tiny world of Mid East specialists.

I came to realize that the field of Middle East studies had been marginalized by political science, in ways that other area studies fields had not. Some of the most influential recent studies of democratization - including projects led by O'Donnell, Schmitter, and Whitehead, Diamond, Linz, and Lipset, and Inglehart - have avoided the Middle East entirely. Even the methodologically-astute scholarship of Pzreworski, Alvarez, Cheibub, and

Limongi simply drops six oil-rich Middle East countries from the database, as though these states occupied a different universe with a unique set of physical laws. In most of these cases the methodological justification for ignoring the Middle East was weak.

A second advantage of this approach is that the results of any tests are likely to be interesting even if they are negative. When we test hypotheses that are original and generated through a deductive process, negative findings are hard to publish: after all, why should others care about our skillfully-generated hypotheses if they are wrong? Testing hypotheses that have already been produced and explored by others, however, is intrinsically interesting: either falsifying or confirming them constitutes scientific progress.

One excellent example is José Cheibub's article "Political Regimes and the Extractive Capacity of Governments" (*World Politics*, April 1998). Cheibub's article closely scrutinizes a common argument: that a government's regime type influences its ability to raise taxes. He shows that theoretical arguments on regime types and taxes - largely produced by Latin America specialists - are often contradictory: some suggest that democracies will extract more taxes, others that authoritarian governments will extract more taxes. His own statistical analysis, based on 108 countries between 1970 and 1990, offers a true out-of-sample test of competing hypotheses. It finds that regime type has no influence on a government's extractive capacity. Even though this is a negative finding, it is still of broad interest because it rejects untested hypotheses that had been widely circulated.

A third advantage of this type of approach is that one can draw on a wealth of empirical insight when it is time to specify one's model. My article drew upon a long series of insightful case studies and theoretical innovations by Eva Bellin, Kiren Chaudhry, Jill Crystal, Terry Lynn Karl, Dirk Vandewalle, and at least a dozen others. Many had developed and illustrated hypotheses about causal mechanisms that linked oil to authoritarianism. I was able to specify and test the three mechanisms that were most widely-discussed.

Of course, understanding these case studies also entailed a lot of work. Often these scholars focused on oil's impact on a variety of dependent variables simultaneously, including regime type, economic performance, and state formation. I had to extract the oil-impedes-democracy argument from a series of wider claims about oil's political effects. Another problem was that there was no consensus among scholars on how the key variables should be defined, how different causal mechanisms operated, and to what domain of cases the oil-impedes-democracy claim should apply. The original concept - which was never terribly precise - had grown fuzzier over time, and sometimes teetered on the edge of being unfalsifiable.

One reason that the oil-impedes-democracy claim had grown stale was that there was little variation on either oil wealth or regime type inside the Middle East region. Another was that almost all Middle Eastern states face at least two other common obstacles to democracy - Islam, and a tortuous colonial history - making it hard to disentangle the partial effects of oil wealth. Together, these factors made the Mid East (ironically) an unusually poor setting for drawing inferences about natural endowments and regime types.

A global data set offered much more variation, and hence, a much stronger test of the claim; it also provided a true out-of-sample test. Since I wished to speak to both Mid East specialists and scholars of comparative democratization, it was important for me to correctly specify my model - that is, to find ways to statistically account for the factors that, according to case studies, had influenced the regime types of oil-rich states. By using a variable that measured the fraction of the Muslim population, I was able to control for the effects of Islam; by using dummy variables for the Persian Gulf states, and for the Middle East and North Africa, I was able to control for any region-specific effects, such as colonial history; and by placing a lagged dependent variable on the right-hand side of the equation, I was able to control for country-specific historical and cultural influences. Even with these and other controls in place, oil significantly reduced the likelihood of democracy.

Comparative politics is hip-deep in causal arguments that have never been subjected to out-of-sample tests. Many have been produced by area specialists through an "inductive" process - which often entails careful field work, historical knowledge, and innovative theorizing. Scrutinizing and testing some of these arguments is not only a good research strategy, it can also foster dialogue between qualitative and quantitative approaches, and area studies and trans-regional approaches to comparative politics.

Notes

[1] I wish to thank Dan Posner for suggesting this.

Nested Analysis in Cross-National Research

Evan Lieberman
Princeton University
esl@princeton.edu

The increasing availability of cross-national statistical data has opened the door for scholars of comparative politics to integrate "large-N" analysis (LNA) with intensive case study analysis to an extent that was simply not feasible in previous generations. Among the central advantages of small-N analysis (SNA) are the ability to measure complex concepts due to close familiarity with the cases and context, and the insights gained from the tracing of over-time processes, in which scholars demonstrate the links between causes and effects. Cross-national statistical analysis provides highly complementary insights through comparisons of much larger numbers of cases, useful for motivating and/or justifying case selection, and for addressing rival explanations, and non-rival control variables. This article describes how to implement a Nested Research Design[1] - which combines the two forms of analysis in ways such that the combination yields greater returns than the sum of the parts.

Nested analysis proceeds by regressing the dependent variable on the relevant and available explanatory variables of interest. If virtually none of the cross-national variation can be explained using existing variables, at the very least, the LNA will demonstrate that certain alternative hypotheses have been systematically considered. Assuming that a relatively good-fitting model can be estimated, however, it is a straightforward task with most statistical software packages to compare "predicted" scores (scores derived using actual values of explanatory variables for a given case in relation to regression parameter estimates) with "actual" scores on the dependent variable, and to create a visual plot of this comparison. Depending upon how one interprets the results of these LNAs, one may elect to carry out either: "on-the-line" case studies, "off-the-line" case studies, or structured comparisons.

"On-the-line" Case Studies

The in-depth component of the nested research design will involve the presentation of cases that "elaborate" the findings from the statistical analysis, after one believes that the regression estimate has done a sufficient job of explaining all of the variance that can reasonably be explained by it. Country cases that are on, or close to, the 45-degree line should be identified, and the link from statistically significant cause to effect can be traced using proper names, actors, events, and historical narrative (George 1979: 46). SNA provides a check for spurious correlation, and can help to fine-tune a theoretical argument by elaborating causal mechanisms. Even skeptics of the inferential power of SNA view it as complementary to LNA: "Case studies are an important complement to both theory-building and statistical investigations… they allow a close examination of historical sequences in the search for causal processes… Comparison of historical cases to theoretical predictions provides a sense of whether the theoretical story is compelling…" (Achen and Snidal 1989: 168-9). Even if case-based knowledge and SNA more generally provided the bases for the LNA statistical tests, scholars may opt to present SNA *after* the LNA for heuristic reasons, as the LNA provides a parsimonious elaboration of the theory, and SNA offers context and detail. Regardless of how they *present* the analysis, scholars should report the sequencing of their analyses.

Both Swank (2002) and Martin (1992) provide examples of book-length studies in which early chapters report statistical analyses that pave the way for SNA. According to Martin (1992: 92), "For those variables that showed statistically significant effects, the analyses complement the case studies by improving our confidence in the generalizability of our results." In each case, LNA provides initial confirmations of their core hypotheses, and dismisses several rival hypotheses, but in both cases, the authors acknowledge that questions about causality arise: Not all of their statistical tests are entirely conclusive, and they acknowledge that a range of possible mechanisms could connect to one another. As a result, they both select cases based on different scores on the central hypothesized explanatory variables, and demonstrate the plausibility of their hypotheses by tracing the impact of alternative scores on those variables to predicted outcomes in the respective cases. Both scholars are deliberate in this approach: "This quantitative work allows me to further refine these hypotheses and provide a framework for the case studies that follow…"

(Martin 1992:11). Both report additional findings and nuances about the cases they describe beyond demonstrating the plausibility of hypothesized relationships from the statistical results. Martin especially highlights contextual factors that could potentially be relevant as more general rival explanatory factors in future research, suggesting a highly transparent methodology.

In the most straightforward manner, nested analysis helps connect explanatory variables to outcomes in ways that might otherwise seem mysterious when statistical results are presented in isolation. For example, Swank points out that large-scale variables such as "international capital mobility" are connected to discrete policy outcomes such as social expenditure through specific historical episodes. Within the case studies, we observe how actors behave, and we are presented with a more transparent accounting of specific policies that are produced within the polity, rather than relying solely upon macroeconomic indicators to serve as proxies for political action.

"Off-the-line" Case Studies

A second variant of the nested research design involves in-depth analysis of cases that are not well predicted by LNA. Indeed, interesting puzzles that are worth studying almost always involve outcomes that defy our "prior" expectations. For example, while there may be disagreement about the precise relationship between democracy and development, we tend to expect wealthier countries to have democratic institutions, and if we were to identify a wealthy country that suddenly became authoritarian this would be surprising and worthy of investigation. However, in many research areas, prior expectations are "non-obvious," and there may be multiple factors that we would expect to influence outcomes. To what extent do these factors really provide sufficient explanations, and how do they combine to influence specific outcomes? In many cases, LNA can provide a baseline estimation of these prior expectations, providing a foundation for a nested analysis by motivating case studies that are "off the regression line." Specifically, this involves study of cases that are distant (typically beyond the 95 percent confidence interval) from the 45-degree line of the regression-predicted vs. actual case score plot described earlier. What is the missing variable, variables, or correct specification of relationships between variables that produces such exceptional outcomes? These questions motivate theorizing and inductive analysis. While the "orthodox" statistically-oriented scholar may describe this as "omitted-variable bias," a more pragmatic perspective suggests that such a technique can help to locate SNA within a scholarly literature and set of debates, and provide an important heuristic device for demonstrating the analytic value-added of that work.

This technique can be used both to motivate the selection of cases from a static perspective, and to motivate puzzles about over-time processes. The most straightforward technique is to pursue outliers from a cross-sectional LNA. When confronted with a fair-to-good-fitting model that explains variance across most cases, a few alternative cases may not fit within the confidence interval, and remain as outliers. The question is, should such cases be regarded as "noise," or as potentially interesting case studies? Given the finite population of country cases with which we have to work, it would be hard to justify neglecting them. When confronted with such questions, scholars need to establish whether they have a hunch about why a given case is an outlier, whether there is a theoretically interesting story to pursue, and whether they have the tools to pursue such questions. In these ways, the LNA provides a more analytically defensible strategy for case selection than the mere statement of intuition, even if it was intuition that motivated the LNA in the first place.

Depending upon the nature of the questions involved, a more dynamic perspective can also be applied. For example, in his methodologically self-conscious study of Venezuelan regime change, Coppedge (2001) motivates the question of patterns of regime change over time in Venezuela through various engagements with theory and conventional wisdom. On the one hand, strictly from the perspective of level of economic development, over-time change in Venezuela would do a relatively good job of predicting this outcome, but as Coppedge explains, other factors are needed, because economic development by itself does not powerfully explain regime outcomes across a larger sample of approximately 4,000 country-years. A case study is useful for identifying additional factors that magnified the impact of economic changes in Venezuela. As Coppedge points out, "The large-sample study identifies what was likely to happen, and then the case study develops deterministic arguments that move us from what was likely (or unlikely) to happen, to what had to happen" (Coppedge 2001:3). From this perspective, the need for the case study is clear: Our existing wisdom on the subject could not account for an important political outcome, and there is room for a new hypothesis or set of hypotheses to help address this conundrum.

Structured Comparisons

Beyond single cases and over-time studies of cases, LNA can be used to motivate structured comparisons for SNA within the larger nested research project, using a mix of "on-the-line" and/or "off-the-line" cases. Statistical relationships between cases can be assessed, and puzzles between cases can be motivated or justified with respect to the larger population of cases through the actual vs. predicted score plot.

The most straightforward example is when countries that would ordinarily be predicted to have "similar" outcomes wind up with different outcomes, perhaps on opposite sides of the regression line, and with at least one case outside the confidence interval. As in the case of "off-the-line" nested analyses discussed above, the role of SNA is to explain variance not accounted for by other theories, but in the case of "structured comparisons," this is with respect to a small set of cases. For instance, in my comparative study of the rise of the tax state in Brazil and South Africa (Lieberman 2003), economic factors alone could explain almost 40 percent of the variance in income tax collections across approximately 70 countries. However, such factors predicted that Brazil and South Africa would have almost identical levels of collection, when in fact South Africa collected about three times as much income tax as Brazil. Such findings helped to motivate a structured comparative analysis of the two countries, in which the goal was to explain this difference.

The use of this technique could help to expand the scope of structured focused comparisons. A great deal of comparative analysis in political science tends to deploy Mill's method of difference to gain analytic leverage. Unfortunately, the requirement of "similar cases" tends to limit scholars to comparing cases within regions, forcing certain sets of comparisons to re-emerge: "France/Germany," "U.S./Canada," "Brazil/Argentina," etc. To a large extent, the underlying logic of such comparisons requires that the scholar make the implausible argument that the two or more countries are "virtually identical" in every way except on the relevant independent and dependent variables. As it stands, the method virtually precludes making comparisons of countries at different levels of economic development because that factor is typically assumed to have a causal influence on virtually all outcomes of interest.

LNA within a nested analysis can help to address such problems, providing a sounder basis for structured comparisons, and opening up the field of possible cross-country studies. To the extent that a measure of the dependent variable can be regressed on several control variables, one can specify prior expectations about similarity and difference, opening up new types of comparisons. For example, comparisons between the U.S. and India might ordinarily be dismissed because of differences in levels of economic development. However, one might find that in a LNA, indicators of development do not hold any explanatory weight, but that colonial legacy (Anglo in both cases), and state structure (federal in both cases) were important predictors of the outcome, leading to similar point estimates. Selecting cases based on point estimates is tantamount to selecting on the *independent* variable(s) of interest. If scores on the *actual* dependent variable turn out to be similar, one could use SNA to study if and how the mechanisms by which federalism and colonial legacy influenced the outcome. Alternatively, if the actual scores were different, one would have a puzzle about the missing variable(s) and/or unspecified interactive relationships. Finally, in a strategy that approximates Mill's method of agreement, one might select cases with *different* regression predicted values, and attempt to explain similarities in actual outcomes. In either case, LNA can set the stage for a comparative analysis that might otherwise seem implausible. The intensive research can then focus much more on demonstrating how certain factors influence outcomes by tracing similarities and differences across countries, and focus less on addressing rival explanations, for which the SNA has limited analytic power.

Iterations of Analysis

Nested analysis may "begin" and "end" in a variety of ways. Research begins when scholars stumble upon questions about cases or patterns that they can explore with available data. One could begin with LNA, move to case study research, and return to the regression. That is, the findings of the case studies could be used to develop new indicators for the relevant explanatory variable and the generalizability of the results could be explored more widely.[2] Over the long term, big research questions could be explored back and forth between large-sample and intensive case analysis across several projects, refining concepts and measures, developing new hypotheses, responding to new conjectures, and gathering new data. The accumulation of knowledge with respect to general questions and questions about specific cases can be greatly facilitated through nested analysis, which invites scholars from both traditions to make contributions to the findings of one another. This view of research is cyclical and dynamic.

When reporting the results of nested analyses it may make good sense to present the statistical and in-depth analyses sequentially. However, the execution of the analysis is likely to involve a constant back-and-forth process. Those who are working on gathering data on a particular case would do well to have a "large-N" dataset handy, just as those who are engaged in statistical analysis would benefit from access to case materials. There are benefits for measurement validity (Adcock and Collier 2001) and for causal inference.

Despite these perceived advantages, there are clearly real and perceived costs of integrating LNA into SNA. Perhaps most importantly, this seems to imply a substantial addition of work. Is this really "two" projects in one? To an extent, yes, but access to cross-national data and statistical software have greatly facilitated the ease with which LNA can be carried out. For example, in the area of democratization research, the Freedom House, Polity, and ACLP datasets provide time-varying indicators across a large number of countries. Similarly, for students of the political economy of development, the World Bank, the OECD, and the IMF publish time-varying economic and other data across most countries for several decades.

The nested research design provides a promising avenue for future cross-national research in comparative politics. The standard procedural and inferential pitfalls of the respective approaches still apply, but the transparent and systematic combining of the two approaches is likely to lead to greater insights about the determinants of important similarities and differences across countries, and to more robust theories of political life.

Notes

[1] Coppedge (2001) provides a very useful, but more narrow specification of the nested research design.
[2] A strategy I pursue within narrowly defined parameters (Lieberman 2003).

Bibliography online on the *APSA-CP* website

Quantitative Empirical Methods and Context Conditionality

Robert Franzese
University of Michigan,
Ann Arbor
franzese@umich.edu

Many, if not most, classical and contemporary comparative political theories either assert or assume that social, political, and/or economic institutional, structural, and/or strategic context modifies causal relationships. For example, culture matters, if it does, by modifying the relationships between objective conditions like poverty and underdevelopment, and outcomes like democratic stability. Individuals' interpretation of poverty and appropriate responses thereto - so a cultural argument may contend- hinge on cultural symbols and understandings. Likewise, institutions matter, if they do, by altering the relationships between objective interests and the institutionally shaped actions perceived as possible and most effective by individuals or groups with those interests. For example, the degree to which some polity's cleavage structure will induce leaders to form political parties based on the societal groups drawn by that structure and voters to support such parties depends on the electoral rules and party-systemic strategic structure that determine the relationships between votes and representation and between representation and governmental power. Complex, context-conditional hypotheses of this sort are now the hallmark of positive comparative politics: the effect of X (e.g., institutions) on Y (outcomes) *depends on* Z (culture, structure, etc.); formally, $dY/dX=f(Z)$. Empirical work, such as described above, which establishes that institutions matter in addition to culture and structure (and *vice versa*) by controlling for the latter in regressions of outcomes on the former, do not reflect this conditionality. They show only that the effect of X (institutions) on Y (outcomes), given or controlling for Z (culture, structure, etc.) is not zero, not that the effect of X on Y depends on Z: formally, $dY/dX|Z=0$ not (necessarily) $\equiv$ $dY/dX=f(Z)$.

Critics of statistical analysis in comparative politics often cite this concern (*inter alia*) that regression coefficients impose a *constant* effect for each independent variable, albeit controlling for others, not effects that differ depending on context (Ragin 1989). That is, they object that broad statistical comparison neglects the context conditionality of comparative politics. This criticism, however, applies only to the simplest linear-additive regression. Multiple statistical devices exist to incorporate the context-conditionality of comparative phenomena (and other complexities) into practical models for empirical analyses. The statistical device most frequently used to evaluate theoretical claims of this sort -that the effect(s) on some dependent variable(s), Y, of some independent variable(s), X, depend upon or are moderated by a third (set of) independent variable(s), Z- is the linear-interactive, or multiplicative, term. In the simplest case, one includes X *times* Z among the regressors. Such interaction terms are hardly new to political science. Indeed, their use is now almost common, yet, especially given current and growing attention to the roles of institutions in comparative politics, they should perhaps become more common still.

As Table 1 shows, 54% of articles since 1996 in leading political science journals use some statistical methods; 24% of those employ interactions. In the exclusively comparative journals, the figures are 49%/25% for *Comparative Political Studies* and 9%/8% in *Comparative Politics*. All of the other journals, except perhaps the two quarterlies, have significant comparative publications, and statistical analyses comprise from 25-80% of those articles, with interactive analyses representing a relatively fixed 20-25%. Thus, about half of top-journal political science articles employ some statistical methods, and about one-quarter of those and over one-eighth of all articles use interaction terms (moreover, these shares include formal-mathematical and philosophical political theory in the denominator). Comparative politics, at least judging by CPS, IO, and WP, operates somewhere between the discipline's mean and half that on these dimensions. The trends in comparative politics and the broader discipline are likely mildly upward in both regards, although these last six years of data suggest a plateau being reached.

Table 1: Types of Articles in Major Political Science Journals, 1996-2001

Journal (1996-2001)	*Total Articles*	*Statistical Analysis*		*Interaction-term Usage*		
		Count	*% of Tot*	*Count*	*% of Tot*	*% of Stat*
American Political Science Review	279	274	77%	69	19%	25%
American Journal of Political Science	355	155	55%	47	17%	30%
Comparative Politics	130	12	9%	1	1%	8%
Comparative Political Studies	189	92	49%	23	12%	25%
International Organization	170	43	25%	9	5%	21%
International Studies Quarterly	173	70	40%	10	6%	14%
Journal of Politics	284	226	80%	55	19%	24%
Legislative Studies Quarterly	157	104	66%	19	12%	18%
World Politics	116	28	24%	6	5%	25%
TOTALS	**2446**	**1323**	**54%**	**311**	**13%**	**24%**

This widespread and perhaps expanding usage of interactions notwithstanding, still more empirical work should contain interactions than currently do, given the substance of many comparative-politics arguments. Consider, for example, the gist of most institutional arguments. In one influential statement of the approach, Hall states, "With an institutionalist model we can see policy as more than the sum of countervailing pressure from social groups. That pressure is *mediated by* an organizational [i.e., institutional] dynamic…" (1986, p. 19; emphasis added).

Thus, in this approach to institutional analysis, and, indeed, inherently in all approaches, institutions are intervening variables that *funnel*, *mediate*, or otherwise *shape* the political processes that translate the societal structure of interests into effective political pressures, and/or those pressures into public-policymaking responses, and/or those policies into outcomes. Extending the list of synonyms might prove a useful means of identifying interactive arguments. When one says X *alters*, *modifies*, *magnifies*, *augments*, *increases*, *moderates*, *dampens*, *diminishes*, *reduces*, *etc*. some effect (of Z) on Y, one has offered an interactive argument. For example, one prominent line of research connects the societal structure of interests to effective political pressure through institutional features of the electoral system: plurality-majority versus proportional representation, etc. (e.g., Cox 1997; Lijphart 1994). Another emphasizes how governmental institutions, especially the number and polarization of key policymakers (veto actors) that comprise it, shape policymaking responses to such pressures (e.g., Tsebelis 2002). A third stresses how the institutional configuration of the economy, such as the coordination of wage-price bargaining, shapes the effect of certain policies, such as monetary policy (e.g., Franzese 2002b:ch.4). In every case, and at each step of the analysis from interest structure to outcomes (and back), the role of institutions is to mediate, shape, structure, or condition the impact of some other variable(s) on the dependent variable of interest. I.e., institutional arguments are inherently interactive, yet, with relatively rare exceptions-see, Ordeshook and Shvetsova 1994, Franzese 2002b, regarding the above cases-empirical work on institutional arguments has ignored this interactivity.

Another example further illustrates the ubiquity of the interactive implications of comparative-institutional theories. Scholars consider principal-agent (i.e., delegation) situations interesting, problematic, and worthy of study because, if each had full control, agents would determine policy, y1, in response to some (set of) factor(s), X, according to some function, y1=f(X), whereas principals would respond to some perhaps different (set of) factor(s), Z, perhaps differently according to, y2=g(Z). Theorists then offer some arguments about how institutional and other environmental conditions determine the monitoring, enforcement, and other costs, C, that principals must incur to induce agents to enact g(Z) instead of f(X). In such situations, realized policy, y, will usually be given by some y=k(C)·f(X)+[1-k(C)]·g(Z) with 0<=k(C)<=1 and k(C) weakly increasing, where the function k(·) describes how contextual conditions (monitoring, enforcement, and transaction costs, etc.), C, dampen principals' control over agents and enhance agents' autonomy from principals. Thus, the effects on y of each c $\in$ C generally depend on X and Z, and those of each x $\in$ X and z $\in$ Z generally depend on C. That is, the effect on y of everything that contributes to monitoring and enforcement costs generally depends on all factors to which the principals and agents would respond differently, and, vice versa, the effect on y of all such factors depends on everything that affects monitoring and enforcement costs. Empirical applications of principal-agent models usually seem to have missed this point.

A rough quantification of the magnitude of such principal-agent and other institutional-interactions omissions from empirical specifications is startling. Of Table 1's 1012 articles with non-interactive statistical analyses, half or so offer some sort of institutional argument. Even if only half of all institutional arguments reflect the interactivity I argue is actually inherent to institutions, that alone would imply that almost as many articles, ½·½·1012=253, incorrectly employ non-interactive empirical techniques to evaluate interactive hypotheses as actually employ interactive terms (311). If, as I expect is closer the truth, institutional arguments are all interactive and many other arguments (e.g., contextual effects in cultural-behavioral theories), say half of those remaining, are also, that would imply that roughly two-and-a-half times as many articles made interactive arguments but empirically evaluated them non-interactively ({½+¼}·1012=759) as actually employed interactions.

The theoretical and substantive interest in such complex, context-conditionality is readily apparent in comparative political economy. As I suggested in a recent review, venerable electoral- and partisan-cycles may be due a theoretical and empirical revisit to explore the institutional, structural, and strategic conditionality of such cycles. For example, in small, open economies, domestic policymakers may retain less autonomy over some policies, or some policies may be less economically effective, so that electoral and partisan cycles in those policies and outcomes are less pronounced than in larger, less-exposed economies. Some polities, moreover, concentrate policymaking control in fewer, more-disciplined partisan actors, which may induce sharper cycles in, e.g., Westminsterian systems than in other democracies. Furthermore, some policies may have more effect and so be more useful and so more used for electoral or partisan purposes, and this too varies with institutional, structural, or strategic context. For instance, the political benefits of demographic versus geographic targeting of spending may vary by electoral system, e.g., single-member plurality favoring the latter and proportional representation the former. These and other contextual variations condition policymakers' incentives and abilities to manipulate policies and outcomes for electoral and partisan gain, and modify the political and economic efficacy of such manipulation, in manifold ways across democracies, elections, and policies, all of which suggests exciting opportunities for interactive models that inform comparative politics. Another obvious locus of interactive effects lies in recent work on *Varieties of Capitalism* (Hall and Soskice 2001) or that on globalization. Still another is the comparative political-economy approach that stresses that the domestic response to international economic integration varies, depending critically on domestic political and institutional context (e.g., Boix 1998, Garrett 1998, Swank 2002). Similar examples from outside political economy are not hard to imagine. The propensity for (apparent) directional voting versus proximity voting in individual electoral behavior, for example, depends on electoral and party systems and the types of government they tend to produce (see, e.g., Kedar 2002).

With so many opportunities to explore interactions-indeed, with the logically inherent interactive nature of comparative political theory-the good news is that quantitative empirical modeling of such context-conditionality can be quite simple. First, one must understand empirical models that embody interactive hypotheses. For example, one typical theoretical argument might be that X generally reduces Y and does so more in the presence of, or the larger is, Z. Note that this is actually two hypotheses: (a) that dY/dX is negative (X reduces Y) and (b) that d2Y/dXdZ is negative (and increasingly so with Z). In a model containing regressors X, Z, and X•Z, such as Y=…a•X+b•Z+c•X•Z…, the interpretation of the results regarding (b) is straightforward. d2Y/dXdZ=c, so the coefficient c simply and directly tells us how the effect of X changes per unit increase in Z and, vice versa, how the effect of Z changes per unit increase in X. (These converses are logically identical; this is not some unique function of regression modeling.) Thus, the standard t-test on coefficient c corresponds to hypothesis (b). The effect on Y of X, dY/dX, however, is not simply a, nor is the effect on Y of Z, dY/dZ, equal to b; nor, even, are these "main" effects of X or Z. The effect on Y of X, dY/dX = a+c•Z depends, as the hypothesis said, on the value of Z (and *vice versa*: dY/dZ=b+c•X). The effects of X and Z each depend on the other variable's value, and the coefficients a or b are just the effects of an increase in X or Z when the other variable equals zero (which need not be "main" in any way, and could even be out-of-sample or logically impossible). In interactive models, the *effects* of variables involved in interactions do not correspond directly to just one *coefficient*; the effects depend on the values of other variables, exactly as argued. Nor do the standard errors (or t-tests) of these *effects* correspond directly to those of any one coefficient; just as the effects of X and Z vary depending on the value of the other, so too do the standard errors of those effects. The best approach for researchers presenting interactive results is to graph or tabulate the *effect* of each variable involved in an interaction as a function of the others, along with the standard errors or confidence intervals of those effects.

Some note, correctly, that the empirical task of distinguishing not just a single, constant effect for X, but one that varies (albeit only linearly) depending on Z, imposes a much heavier burden on the data. This is also the substantive meaning of concerns expressed regarding the high multicollinearity (i.e., correlation) among regressors X or Z and X•Z in interactive models. Moreover, one must discard the notion that "centering" the interacting variables (subtracting their means), as several methodological texts advise, eases this empirical task. *Centering alters nothing important mathematically and nothing at all substantively*.[1] The multicollinearity concern is quite valid, then. The empirical task that interactive analyses set is very demanding, and these demands will heighten dramatically as the number of interactions increases, as the complex, context-conditionality of comparative politics suggests they should. However, this concern, too, is an unavoidable *logical necessity*. Comparative researchers have four options, each with characteristic perils:

1) Ignore the context-conditionality of their arguments by omitting interactive terms. Judging by Table 1, most analysts do this, but this does violence to the inherently (and interestingly) interactive nature of comparative politics and plagues those effects actually estimated with omitted-variable bias and statistical inefficiency.

2) Reduce context-conditionality by allowing only one or few of the hypothesized interactions in their model. This enables more exclusive focus on those included interactions and reduces the omitted-variable biases and inefficiencies relative to excluding interactions altogether, but it does not eliminate these problems and it ignores the complexity of the context-conditionality in comparative politics.

3) Constrain the context-conditionality to follow a specific functional form suggested by theory. This reduces the empirical demand on the data in revealing more of the theorized complex, context-conditionality in comparative phenomena, thereby reducing still further the misspecification and inefficiency problems of the above two approaches. However, many comparative theories may not be sufficiently precise to determine the form of interactions. The gained strength arises from leaning more heavily on theory, and the multicollinearity concerns re-emerge as the allowed complexity increases, albeit at a lesser pace.

4) Conduct qualitative empirical analysis to supplement or substitute for quantitative analysis. This may partially counteract the lack of information underlying the multicollinearity problem by enriching the detail and depth of the empirics, but it typically enhances the quality of the information thusly at the cost of severely reducing the quantity (see, e.g., King

et al. 1994 for advice on making such trades optimally) and the ability to discern complex interactions qualitatively. I.e., without precise numerical measurement and statistical control, determining interactive effects of combinations of independent variables becomes inherently more difficult.

I advocate the third option. Because the problems raised by complex context-conditionality are logically inherent, qualitative analysis cannot evade them. The other two options evade them only to the degree that they suppress the (interesting) conditionality. To see the promise of the third approach, return to the principal-agent (i.e., delegation) case described above. Generally, in such situations, we argue that, if each had full control, agents would act according to some function, $y1=f(X)$, while principals would act differently, $y2=g(Z)$. We then argue that some institutional and other contextual conditions determine the monitoring, enforcement, and other costs, C, principals must incur to force agents to enact g(Z) instead of f(X). Realized policy, y, will then typically be given by some $y=k(C)\cdot f(X)+[1-k(C)]\cdot g(Z)$ with $0<=k(C)<=1$ and k(C) weakly increasing as noted. If the comparative theory can identify k(C), the functions $f(\cdot)$ and $g(\cdot)$, and factors X and Z that state to what and how principal and agent would respond if left (hypothetically) completely in charge, and if these functions and/or factors are not identical, then non-linear regression techniques can gain leverage on *all* the complex conditional effects predicted in that comparative context. Moving beyond delegation to other shared policy-control situations, researchers might also fruitfully apply this approach to study the relative weight in policy control of, e.g., executive and legislative branches in (semi-) presidential systems, or of different chambers in multicameral systems, or of prime-, cabinet-average-, cabinet-median-, and portfolio-ministers in parliamentary systems, or of committees or cabinets and legislature floors or backbenchers or oppositions, or, even, of the degree to which elected representatives act legislatively as if they represent the residents of their electoral district, those therein who support them, or their national-party's or some other constituency. Finally, even more generally, researchers can apply similar non-linear approaches to any situation in which some factor or set of factors modify the impact of several others proportionately, thereby bringing many more of their highly interactive theoretical propositions under empirical scrutiny than perhaps previously thought possible. Indeed, institutions often operate in this way. For example, institutions that foster greater party discipline may induce legislators to behave less (geographically) distributively and more (class/ideological) redistributively, implying a proportionate modification in their response to a range of political economic conditions. Similarly, institutions that facilitate voter participation tend to broaden the distribution of interests represented in the electorate and so therefore influencing policy, again suggesting that electoral institutions will modify the effect of many political-economic conditions on government policies *proportionately*.

The approach is not panacea, of course. It does require that researchers know well how policy would be determined under the hypothetical extremes of key parameters and that the inputs to these policy-response functions vary empirically in sample, and it gains empirical leverage and produces truly revealing estimates of those parameters only to the degrees they actually do so. Still, many important substantive problems in comparative politics, and in positive political science more generally, involve similar complex, context-conditional relationships, and this approach seems to offer a more theoretically, methodologically, and empirically promising way to address those issues than the alternatives.

Notes

1 Likewise, the oft-raised concern that multiplicative terms cannot distinguish, for example, $X \cdot Z=12$ with $X=3$ and $Z=4$ from $X \cdot Z=12$ with $X=2$ and $Z=6$, is of little import because the model, that is the model of the effect of X and Z on Y, can and will distinguish those cases insofar as they actually do differ logically. Incidentally, the common admonition that one should include both X and Z if the model contains an $X \cdot Z$ term is misleading. It is usually a highly advisable philosophy-of-science guideline (a mathematical application of Occam's razor) but is neither a logical nor a statistical necessity.

Bibliography available online on the *APSA-CP* website

Islam and Comparative Politics

Lisa Anderson
Columbia University
la8@columbia.edu

John Esposito, *Unholy War: Terror in the Name of Islam* (New York: Oxford University Press, 2002)

John Esposito and John Voll, *Makers of Contemporary Islam* (New York: Oxford University Press 2001).

Gilles Kepel, *Jihad: The Trail of Political Islam*, translated by Anthony F. Roberts, (Cambridge, MA: The Belknap Press of Harvard University Press, 2002)

Bernard Lewis, *What Went Wrong? Western Impact and Middle Eastern Response* (New York: Oxford University Press, 2002)

Fatima Mernissi, *Islam and Democracy: Fear of the Modern World*, second edition, translated by Mary Jo Lakeland, (Cambridge, MA: Perseus Publishing, 2002)

Olivier Roy, *The Failure of Political Islam*, translated by Carol Volk, (Cambridge, MA: Harvard University Press, 1994)

"What Went Wrong?" - this is the title of the best known of the books under review and it is designed to convey an interpretation of the modern history of the Middle East as one of decline. I will return to this thesis but first I will borrow the title to explore an odd feature of the field which all these books are meant to represent: comparative politics of the Middle East.

None of the authors whose work this review discusses are card-carrying American political scientists, and none of them pay any attention, much less deference, to the conventions of American-style comparative politics. There is no ritual discussion of "the literature" and there is no effort to deploy this empirical material to illuminate political science puzzles. Gilles Kepel is a political scientist and Olivier Roy is a sociologist but both work in a European tradition much less concerned with the "science" of politics than their American counterparts; Fatima Mernissi is a Moroccan professor and activist, trained as a sociologist in the United States but little concerned with the methodologies or debates that seize her American co-disciplinists. Bernard Lewis and John Voll are historians; John Esposito has usually taught in religion departments.

None of this makes them bad scholars, of course, but it does raise the question of where the comparative political scientists whose work might ordinarily be the subject of a review in this newsletter are. Is this particular array of books a result of selection bias, or is there really something missing in the world? The answer is, in a sense, both. Defining the research terrain as "Islam and politics," does bias the selection. Comparativists in other regional studies do not typically spend a lot of time these days on "religion and politics," being more interested in, say, democratic institutions, economic development, or perhaps ethnic conflict (in which religion may be an identity but is not, typically, an ideology or system of values). Not only does the study of Islam limit the empirical reach of research to the Muslim world (as opposed to all democracies, developing countries or civil wars, for example) but religion as such is simply not very interesting to opinion makers in political science right now.

Had we defined the research terrain as "Middle Eastern politics," however, there would not have been much more to add to the list of books published in the last five years. Comparativists have not been working in the region recently. There are little more than a dozen junior American comparative politics PhDs who conduct research on the Arab world currently employed at any American university or college, much less in those with Title VI National Resource Centers. Double that number if you add scholars of international relations and political sociologists. The Middle East Studies Association lists several hundred political scientists but many of these do not maintain active research agendas; the association is overwhelmingly made up of history and humanities faculty.

Why? In part because it's hard, and getting harder, to conduct field research. Americans are not able to conduct research at all in a number of Middle Eastern countries--Libya, Iraq, Iran, Saudi Arabia--it is dangerous in a number of others--Algeria, Palestine, Afghanistan--and it is difficult nearly everywhere else. Moreover, with the exception of Israel, Turkey and perhaps Indonesia, there is almost no local social science research community with which to collaborate--and those which exist are increasingly uncomfortable working with Americans. Finally, there is no professional audience in the United States for the kind of research problems that make the most sense in the region. While students of the Middle East find they have little occasion to join research projects on the nature of democratic institutions, for example, few comparativists in other regional studies are possessed by the problems of ruling monarchs, by the distortions of labor markets created by pronounced legal gender inequality, by the perversions of globalization precipitated by censorship, etc. (Many of these issues are taken up in the new, and remarkably candid *Arab Human Development Report 2002*--the work of some of the best social scientists in the Arab world.[1])

In the absence of alternatives, we read books about Islam, in the hope that if they will not help us advance our discipline, at least they will help us decipher the front page headlines. From this vantage point, the books under review represent perspectives that diverge in interesting ways. Bernard Lewis' *What Went Wrong? Western Impact and Middle Eastern Response* enjoyed quite a long run on the bestseller lists in the wake of the attacks of September 11th and it provides what can best be described as a short, elegantly written version of the conventional Western wisdom--pretty much what graduate students were taught forty years ago about Islamic history. The West has all the active verbs ("Western impact") and Islamic world all the passive ("Middle Eastern response"). This dynamic is a reflection, we are told, of the natural energy of scientific, economic and political innovation spilling out from Europe. There is no mention of European imperialism, except as one of "scapegoats" Muslims dredge up when trying to account for the dismal state of affairs today. There is no discussion of the reform efforts undertaken by local rulers in face of European competition (except a dismissive reference to several mid-19th century efforts at constitutionalism which contains a minor but, from one of the most distinguished historians of the region, surprising error of fact--the French Protectorate in Tunisia was established in 1881, not 1864) or of current political trends at all.

Many Muslims, including not only Osama bin Laden but most of the intellectuals in the Muslim world, would agree with Lewis that something went wrong, but they would not agree that a story which attributes a benign modernity to the West and a supine obscurantism to the Middle East is adequate to understanding, much less rectifying the situation. Nonetheless, that "what went wrong" is best framed in civilizational terms is a widely shared perspective: Samuel Huntington's *Clash of Civilizations* finds a mirror image in the Muslim world.[2] From this perspective, the problems in the region are not to be explained by the economic and political impact of twentieth century imperialism, nor by the proxy wars of the Cold War superpower rivalry, nor by the impact of oil rents and spending (all potential rival hypotheses) but by cultural and religious hostility.

In this respect, the remarkable insight of Edward Said twenty-five years ago--that the conventional wisdom exemplified by Bernard Lewis (they are leaders of opposing camps in Middle Eastern studies) reflected an unstated bias born of centuries of Orientalist writing about Islam and Muslims--has had a mischievous effect.[3] Many students of the Middle East and of Islam, both within the region and in "the West," have taken the debate about Orientalism to be the essence of the analytical landscape. Literary criticism, however insightful and important, does not exhaust the analytical traditions that may be deployed in understanding the region but the importance of Said's critique had the effect of privileging the cultural domain of intellectual and scholarly endeavor for decades thereafter. The "clash of civilizations" feeds upon this emphasis on culture and, in the absence of significant empirical social science, it finds resonance in the intellectual and policy debates in the region as well as in the US and Europe.

John Esposito has long played as important role in providing a window on relatively temperate Muslim perspectives on this "clash." In *Unholy War: Terror in the Name of Islam*, he takes on the issue directly, arguing that although Huntington's argument hit a responsive cord in the region (there is, he tells us, a "'market for clash,'" p. 126), there is no single or dominant perspective on Islam or on its relationship with the non-Muslim world, especially the West within the Muslim world. In *Makers of Contemporary Islam*, he and his frequent collaborator, John Voll, provide examples of nine "Muslim activist intellectuals" whose work demonstrates this contention. Although there are those who think Esposito favors moderate voices, and therefore distorts the true picture of the dangers Islam may contain, in fact, it also contains much promising intellectual ferment. Just as Samuel Huntington does not summarize American views of the world, or of Islam, there is considerable diversity of opinion in the Muslim world itself. In the current climate of what seems to be mutual assured fanaticism in the popular media, there is certainly nothing wrong with reminding people of this elemental fact, and Esposito's work is accessible and reassuring.
Fatima Mernissi's *Islam and*

Democracy is an idiosyncratic, even eccentric, effort to meld cultural interpretive traditions with more conventional sociology. She agrees with Lewis that, as he puts it, "at the present time secularism is in a bad way in the Middle East," (p, 108), although Mernissi attributes that hostility to secularism to the recent historical need to resist Western imperialism by retreating into culturally authentic Islam. Whatever the reason, this has meant, in Mernissi's estimation, that "the modern Muslim state...has never committed itself to teaching individual initiative...and did not anchor modern identity in the rationalist tradition." (p.42) Hence, political institutions and behaviors that seem self-evidently transparent in the West are incomprehensible to Muslims. This sort of argument, which is not unusual among Muslims trained in the West, represents interesting and ambitious claims that beg for some kind of empirical operationalization. In the absence of any effort to test them, however, these arguments smack of exactly the sort of essentializing impulses for which Edward Said took Western scholars to task.

Gilles Kepel and Olivier Roy have done the most disciplined empirical research of the authors reviewed here, and their works share an interesting, and probably distinctively European, perspective. They both argue that the Islamist opposition movements which have drawn such attention to the role of Islam in politics in the last several decades actually peaked in influence in the late 1980s. For Roy, they are the heirs of the socialist and communist movements that championed the Third World during the Cold War; for Kepel, they are the extension of the anti-imperialist nationalist movements. In either case, the early proponents of Islamic socialist revolution or Islamic states did not succeed--Libya and Iran having probably come closest to these aims--and in the face of declining popular support, the movements themselves radicalized, becoming ever more extreme in their death throes. As Roy puts it, the failure of the Islamist revolutionary idea brought about the drift...toward a puritanical, preaching, populist, conservative, neo-fundamentalism..." (p. 25) The extremists, seeing themselves unable to mobilize mass revolution, resorted to dramatic, symbolically evocative acts of terrorism, both to convey an importance they do not actually enjoy and in the vain hope of reviving the revolutionary consciousness of the masses. In this, they are little different, argues Kepel, than the Red Brigades in the waning days of communism.

At long last, we have a genuine hypothesis that might intrigue a comparative political scientist! It is certainly not one that would have arisen from a reading of the American popular media, with their predilection for essentializing and demonizing Islam, nor from the puzzles routinely posed in the scholarly and scientific journals of American social science, inclined as they are to favor research terrains that are either easy or "in transition" towards something familiar. The existence of an alternative, in this instance French, intellectual tradition provides a salutary reminder of the research opportunities of which American social scientists may be depriving themselves.

If reforming comparative politics to better encompass the Muslim world seems a bit daunting but you still want to read one book on "what's going on" there, choose Kepel. Comprehensive and thorough, well written and translated, this is both an excellent survey of Islamist politics and a intriguing argument. Not only will you learn a lot that will help with deciphering the headlines, but you may just be provoked to include the Middle East and the Muslim world in your comparative lens after all.

Notes

[1] United Nations Development Programme, 2002
[2] The *Clash of Civilizations and the Remaking of the World Order* (New York: Simon & Schuster, 1997)
[3] *Orientalism* (London: Routledge, 1978)

Surveying the Field: Basic Graduate Training in Comparative Politics

Annabella España-Nájera
University of Notre Dame
mespanan@nd.edu

Xavier Márquez
University of Notre Dame
xmarquez@nd.edu

Paul Vasquez
University of Notre Dame
vasquez.11@nd.edu

1. Introduction

This article explores what graduate students in comparative politics are expected to learn in their first few years of training, based on what they read in survey courses or in preparation for their comprehensive exams. Our findings contain some surprises. We found that there is little agreement on the kind of guidance graduate students should receive as they prepare for these exams and little agreement on the specific works that constitute the core thematic readings in the field. However, there is much more agreement on a very small number of works and on the range of topics and approaches that should be mastered, as well as on the regions most often studied.

2. Methodology and Research Design

We solicited comparative politics comprehensive exams (also known as prelims or qualifying exams) reading lists from the National Research Council's list of the top 50 political science departments in the country as they were ranked in 1993. Since several institutions were tied for the 50th position in the NRC rankings, there were in fact 52 departments in our sample. Departments that did not have official reading lists for comprehensive exams were asked to send in a syllabus from their comparative field survey course, if one was offered, so that we would have some data to represent the literature in comparative politics that was emphasized in their graduate program. All institutions were told that information would not be analyzed or communicated in such a way that readers could identify which readings were required by any specific program. Thirty-two of the 52 departments provided this information, yielding seventeen reading lists and fifteen syllabi.[1]

We believe that whether or not departments had reading lists to provide us with tells us something important about different perspectives regarding the training of graduate students and preparation for qualifying exams. Some programs may have reading lists because faculty believe that they owe their students detailed guidance. Conversely, some programs may eschew official reading lists, because they believe that the formulation of a reading list by each student for his or her own exam preparation is an important pedagogical tool or rite of passage. Finally, some departments may emphasize study relying largely on survey course syllabi and class work, because they allow for more current works to be introduced and covered. Depending upon the flexibility in a department's field, these approaches need not be mutually exclusive. Forty-five departments either provided a reading list or a syllabus or indicated that they had no reading list. Of those 45 respondents, 62% did not have an official reading list, which indicates a greater emphasis on student-created reading lists or reliance on introductory course material.

We preferred to use reading lists rather than syllabi for two reasons. First, we believed that reading lists would probably be more representative of collective agreement, since reading lists would be constructed with greater attention to reaching consensus at the field level in each department. Second, we believed that reading lists would be more comparable than syllabi, because the number and diversity of readings as well as the quantity of pages assigned on syllabi would be more constrained if an academic institution was on the quarter rather than the semester system. While some of our analysis relies on inferences from data pooled across reading lists and syllabi, the data that are publicly available from this project will allow scholars to analyze data separately for reading lists and syllabi. In a few circumstances where departments provided us with both reading lists and syllabi, we entered only data from the reading list in light of 1) the rationale explained above and 2) to preserve comparability between institutions. Although some syllabi offered sections of readings that were recommended or suggested, these were not included in our data because not all syllabi made similar distinctions.

From the information provided by the

32 departments, more than 1,050 distinct works were collected. In addition to the title of each work, the author(s), year of publication, if applicable, and source of the work were recorded (i.e., whether it was an entire book or simply a chapter from a book or a journal article).[2] Also, the number of assigned pages was recorded. If the exact number of pages was not specified, we assumed that the entire work was assigned. In the absence of specific page numbers, we assumed that books were 300 pages in length, chapters were 30 pages in length and that journal articles were 20 pages long. Second, each reading was weighted to reflect the degree to which students would be held accountable for being familiar with a given work. In most cases, no distinction was made on the lists and works were given a weight of one. However, if a department required students to prepare for readings on two out of four lists, for example, each reading on the four lists was given a weight of .5, since the two lists for which the student would be accountable would be equivalent to a weight of one. These weights were aggregated in two ways. First, the total weight for a work is a sum of the total weights across departments that assign a book, chapter or article. This number represents the frequency with which departments assign a given work. Second, the total weight of the sample is equal to the sum of the total weight for all works. This number is most useful as the denominator when calculating the percentage of the entire sample that is represented by a given subset.

Finally, we coded readings to reflect their thematic and geographic relevance. For example, if a piece addressed political culture it was given a code associated with political culture. Up to three codes could be assigned to any particular reading. Similarly, as many as two regional codes could be assigned to any individual work.[3] Consequently, categorizations by theme and region allow for a reasonable amount of overlap.[4]

3. Comparing the Amount of Work

Using the data we collected, it was possible to count the number of works that each department assigns in its reading list or field course syllabus as well as an estimate of the total number of pages of reading required. We could thus find the minimum and maximum number of works and pages assigned, as well as the median and mean (see Table 1).

Table 1

	Syllabi	Reading
Mean number of works	54	97
Median number of works	47	81
Mean number of pages	4978	17383
Maximum number of works	116	184
Minimum number of works	9	33
Maximum number of pages	15897	36210
Minimum number of pages	1021	8197
Standard deviation (works)	32.4	48
Standard deviation (pages)	3920.5	8580

As with most of our other findings, there seem to be substantial variations across departments on the amount of work assigned to graduate students. This was perhaps expected. We were nevertheless surprised at how much variation we observed in the distribution, and were unsure as to what would best explain it. Are the differences explained by the fact that some departments are more demanding than others? If this is the case, are there potential consequences for the quality of education that graduate students are receiving across departments? We are aware, of course, that quantity does not always equal quality.

A less bleak explanation is perhaps that some departments provide graduate students with less guidance as they prepare for their comprehensive exams. A substantial number of respondents told us, after all, that their graduate students were expected to develop their own reading lists. This method of developing a reading list also raises some questions, however, regarding how much attention and guidance graduate students are receiving as they work to develop their reading lists. Do all faculty members put the same amount of thought and time into the process? Another potential consequence of this kind of guidance is the perpetuation of disagreements and variation in the field over what is essential reading for all comparative graduate students.

We must note that our attempts to make sense of the sharp differences across departments are all *post hoc*, and certainly they cannot be confirmed with the information that we have available to us. We nevertheless thought it worthwhile to speculate on the different amount of work being assigned across departments, given the sharp differences observed and the potential consequences of these differences.

4. Pooling the Data

As mentioned above, we draw some

of our inferences from data pooled across reading lists and syllabi. We decided to work from a pooled list because when a comparison is made between those works most often assigned in the reading lists and the syllabi, a substantial amount of similarity is found. Both sets of lists tend to assign the same top works, and while some top works might have a higher weight for one of the lists than the other, there are very few top-thirty works on the pooled list that are not also top-thirty works on the unpooled lists.[4] Our reasons for using a pooled list become clear when we observe Table 2, which reports the top thirty works in the pooled list with the weights that the works received in both the reading lists and the syllabi.

Table 2

Author	Short Title	Year	Weight in pooled list	Weight for reading list only	Weight for Syllabus only
Putnam with Leonardi and Nanetti	*Making Democracy Work: Civic Traditions in Modern Italy*	1993	26.5	15.5	11
Huntington	*Political Order in Changing Societies*	1968	24.5	15.5	9
Moore	*Social Origins of Dictatorship and Democracy*	1966	24	15	9
Skocpol	*States and Social Revolutions*	1979	23.5	15.5	8
King, Keohane and Verba	*Designing Social Inquiry*	1994	21	13	8
Almond and Verba	*The Civic Culture*	1963	18.5	11.5	7
Downs	*An Economic Theory of Democracy*	1957	18.5	13.5	5
Dahl	*Polyarchy*	1971	18.5	14.5	4
Bates	*Markets and States in Tropical Africa*	1981	17.5	13.5	4
Olson	*The Logic of Collective Action*	1965	15	11	4
Anderson, B	*Imagined Communities*	1991	14.5	10.5	4
Przeworski	*Democracy and the Market*	1991	13.5	9.5	4
Tarrow	*Power in Movement*	1994	13.5	10.5	3
O'Donnell, Schmitter and Whitehead	*Transitions from Authoritarian Rule: Tentative Conclusions*	1986	13.5	11.5	2
Huntington	*The Third Wave*	1991	13.5	12.5	1
Evans, Rueschmeyer and Skocpol	*Bringing the State Back In*	1985	13	8	5
Przeworski and Teune	*The Logic of Comparative Social Inquiry*	1970	13	9	4
Gerschenkron	*Economic Backwardness in Historical Perspective*	1962	12.5	8.5	4
Lijphart	*Comparative Politics and the Comparative Method*	1971	12	7	5
North, D	*Institutions, Institutional Change and Economic Performance*	1990	12	9	3
Cox	*Making Votes Count*	1997	11.5	9.5	2
Bates et al.	*Analytic Narratives*	1998	11	4	7
Lipset	*Poltical Man*	1960	11	7.5	3.5
Ragin	*The Comparative Method*	1987	11	8	3
Sartori	*Parties and Party Systems*	1978	11	11	0
Hall, P	*Governing the Economy*	1986	10.5	6.5	4
Gellner	*Nations and Nationalism*	1983	10.5	6.5	4
Rueschemeyer, Stephens and Stephens	*Capitalist Development and Democracy*	1992	10.5	6.5	4
Olson	*The Rise and Decline of Nations*	1982	10.5	8.5	2
Cardoso and Falletto	*Dependency and Development in Latin America*	1979	10	5	5

By using this pooled list for our analysis, we do not lose a significant amount of information.

5. Disagreement in the Field: Are There Any "Required Readings"?

Possibly the most striking fact that we uncovered is the degree of disagreement within the field. Only 9 (out of 1066, or less than 1%) works are assigned by more than fifty percent of the departments surveyed, and only 29 by more than a third. There is less disagreement if we only look at the departments that have reading lists (17 out of 32 departments responding), but we still find a great deal of fragmentation: only 22 works (out of 747, or about 3%) are assigned by more than half of the departments with a reading list, and 49 by more than a third. Few works seem to be "required reading" in the field.

To be sure, it is possible that this disagreement is more apparent than real. Many works by different scholars make similar arguments. Different departments may assign different works by the same authors, contributing to a misleading impression of fragmentation. Furthermore, our sample does not fully represent what graduate students are expected to read before taking their comprehensive examinations, since they normally take a variety of courses in which they read far beyond what is included in the particular reading list at their department or in the syllabus of the comparative politics survey course.

We do not have, at the moment, a good way to test whether the diversity, even fragmentation, evidenced by our sample masks a great uniformity of argument. A preliminary factor analysis of the sub-sample of works coded by topic (not reported here) shows that most departments attempt to cover a representative range of approaches to comparative politics. Informal inspection of the range of works listed by each department also suggests that students in almost all departments are expected to become acquainted with a wide variety of approaches, viewpoints, and arguments. Our coding of the data for topic, however, has severe limitations, as already discussed (see note 3), so these observations should be taken with a grain of salt.

Looking at the representation of authors rather than of particular works does diminish a bit the degree of fragmentation. Instead of 1066 works, we find about 672 authors, depending on how collective authors are counted. Of these, 65 authors (about 10%) account for half of the total weight in the sample. No author, however, has his/her works assigned by every responding department, and only 23 authors have their works assigned by more than half of all responding departments. Putnam, Huntington, Skocpol, Barrington Moore, and Lijphart, unsurprisingly, are the top five most-assigned authors (measured by the number of departments that assign at least one of their works). As we saw in Table 2, (some of) their works also occupy top ranks among the most assigned.

6. Most Assigned Works

We describe works as "canonical" here if they are assigned by more than one third of all departments in the sample. Looking more closely at the only nine works (out of the 29 canonical works - see Table 2 above) assigned by more than half of the departments, we found no real surprises. Anthony Downs' *An Economic Theory of Democracy* is the oldest work in the list and the only one that explicitly employs formal theory in its analysis, though Bates' *Markets and States in Tropical Africa* is very much informed by rational choice. The most recent work, King, Keohane and Verba's *Designing Social Inquiry*, is the only work specifically focused on methodological questions. Three of the works were written in the 1960s (four if we count *An Economic Theory of Democracy*, written in 1957), two in the 1970s, one in the 1980s, and two in the 1990s. No work written in the last 8 years has made it to this list, possibly suggesting that some time must pass before there is broad agreement about the value of a given work.

Four of these works are concerned with cross-regional analyses. These are Huntington's *Political Order in Changing Societies*; Skocpol's *States and Social Revolutions*; Almond and Verba's *The Civic Culture*; and Dahl's *Polyarchy*. Of the remaining works, two do not have a regional focus but are more concerned with theoretical and methodological questions (Downs's and King, Keohane and Verba's). The rest have a distinct regional focus. Putnam's *Making Democracy Work* (the most frequently assigned book) and Moore's *Social Origins of Dictatorship and Democracy* focus on Western Europe, while Bates' *Markets and States in Tropical Africa* is about Africa. It is interesting to note that most of them make a strong contribution to our theoretical understanding in the discipline, despite concern with a particular area in some cases. This finding seems to suggest that the criterion for joining this most often assigned group of works is a substantial or unique theoretical contribution to the field.

8. Analysis of Works by Topic

The distribution of works by topic reveals some interesting facts. Works on Political Economy form the largest group in the sample, accounting for 19% of the weight, followed by works on Regimes and Democratization (15% of the weight) and Methodology (13%). (See Table 3). It is unclear why works on Political

Economy should account for so much of the total weight; perhaps there is an unusually large number of such works that address theoretical issues and problems of interest to the discipline, or perhaps issues in political economy are thought to be of great importance today (though issues of Identity Politics and Nationalism, which are arguably very important today too, account for only a meager 5% of the weight of the sample).

As we can see from Table 3, there is often considerable overlap among the canons of various thematic categories, as some works recur again and again. This suggests that a number of topics could possibly be merged. For example, if the categories of Parties and Party systems and Electoral Systems were taken together, they would account for 9% of the total weight (instead of 7% and 6%, respectively, if taken separately), with 89 works among them (instead of 62 and 50 if taken separately). The categories of constitutional design, subnational politics and democratic performance could also be merged with broader categories (Institutions, Political Development or Regimes and Democratization) without changing much in the table.

The category of History and Overview of Comparative Politics has no canonical work - in fact, it has no work assigned by more than four of the responding departments. This is relatively unusual; even in the other categories without a canonical work - International Relations and Constitutional Design - there is far more agreement regarding the works that are worth reading. This suggests that comparativists do not agree either about the history of their own discipline or about the importance of teaching that history.

Table 3

Topic	Weight in Sample[5]	Number	Median Year	Canonical Works[6]
Political Economy	19%	173	1990	*Democracy and the Market* (Przeworski)
				Economic Backwardness in Historical Perspective
				Institutions, Institutional Change and Economic Performance (North)
				Political Man (Lipset)
				Capitalist Development and Democracy (Rueschemeyer, Stephens and Stephens)
				Governing the Economy (Hall)
				The Rise and Decline of Nations (Olson)
Regimes and Democratization	15%	127	1992	*Social Origins of Dictatorship and Democracy* (Moore)**
				Polyarchy (Dahl)*
				The Third Wave (Huntington)
				Democracy and the Market (Przeworski)
				Transitions from Authoritarian Rule (O'Donnell, Schmitter and Whitehead)
				Political Man (Lipset)
				Capitalist Development and Democracy (Rueschemeyer, Stephens and Stephens)
Methodology	13%	116	1991	*Designing Social Inquiry* (King, Keohane, and Verba)*
				Bringing the State Back In (Evans, Rueschemeyer, and Skocpol)
				The Logic of Comparative Social Inquiry (Przeworski and Teune)
				Comparative Politics and the Comparative Method (Lijphart)
				The Comparative Method (Ragin)
				Analytic Narratives (Bates et al.)
States and State-Building	10%	84	1988	*States and Social Revolutions* (Skocpol)**
				Markets and States in Tropical Africa (Bates)*
				Bringing the State Back In (Evans, Rueschemeyer, and Skocpol)
				Economic Backwardness in Historical Perspective
Political Development	8%	67	1979	*Making Democracy Work* (Putnam)**
				Political Order in Changing Societies (Huntington)**
				Social Origins of Dictatorship and Democracy (Moore)**
Formal Theory	8%	58	1994.5	*An Economic Theory of Democracy* (Downs)*
				Markets and States in Tropical Africa (Bates)*
				The Logic of Collective Action (Olson)
				Making Votes Count (Cox)
				Analytic Narratives (Bates)
Political Culture and Mass Behavior	8%	67	1990	*The Civic Culture* (Almond and Verba)*
				Power in Movement (Tarrow)
Parties and Party Systems	7%	62	1990	*An Economic Theory of Democracy* (Downs)*
				Making Votes Count (Cox)
				Parties and Party Systems (Sartori)
Elections and Electoral Systems	6%	50	1990	*An Economic Theory of Democracy* (Downs)*
				Making Votes Count (Cox)
				Parties and Party Systems (Sartori)
Rebellion, Revolution and Violence	6%	42	1985	*Political Order in Changing Societies* (Huntington)**
				States and Social Revolutions (Skocpol)**
Institutions	5%	61	1993	*Institutions, Institutional Change, and Economic Performance* (North)
Interest Groups	5%	53	1991	*The Logic of Collective Action* (Olson)
				The Rise and Decline of Nations (Olson)
Ethnicity, Identity Politics, and Nationalism	5%	54	1991.5	*Imagined Communities* (Anderson)
				Nations and Nationalism (Gellner)
Participation and Collective Action	4%	35	1990	*The Logic of Collective Action* (Olson)
				Power in Movement (Tarrow)
History and Overview of Comparative Politics	3%	45	1993	No canonical work
Democratic Performance	3%	27	1992	*Making Democracy Work* (Putnam)**
Subnational Politics	3%	27	1989	*Making Democracy Work* (Putnam)**
International Relations	3%	34	1992.5	No canonical work
Constitutional Design	2%	11	1990	No canonical work

9. Analysis of Works by Region

Several interesting tendencies emerge when examining the reading lists by the amount of attention devoted to different parts of the world. Works with a cross-regional focus constituted the highest percentage of the total weights in the sample (29%) in our pooled regional data sample. These include both works explicitly classified as cross-regional as well as works dealing with two or more regions. In addition to having works that are often assigned, this category has the largest number of titles (160) when compared with all other regional categories. Of all 160 pieces in the cross-regional category only eighteen readings achieve canonical status. Writings on Western Europe (excluding cross-regional works) are somewhat analogous to those with a cross-regional focus in that they are highly weighted out of the total sample weight (11%) and have a wide number of readings that are assigned (99).

The regional writings with the next-highest weight of the total come from Latin America and the Caribbean (3%) and the United States (2%) respectively, excluding cross-regional works. (If we include cross-regional works in these categories, however, the numbers rise to 5% and 4% respectively). In addition to having similar weighted values, these literatures have comparatively fewer associated titles than the regions discussed previously and their canon is also narrower. The category of works on Sub-Saharan Africa has an even smaller overall weight, list of titles, and canon. It also has few cross-regional works; excluding them from the count of works on Africa does not reduce the weight of the category appreciably.

The remaining regional categories for Eastern Europe, Middle East and North Africa, East Asia, Russia and the former Soviet Republics, Southeast Asia and South Asia all individually constitute one percent of the weighted total percentage. Readings covering these areas have a cumulative weight of only six percent for the total sample and only fifty-three titles at most that address these parts of the world. By contrast, this is equivalent to the weighted values and numbers of works required for study of Latin America or the United States. Consequently, there appears to be a clear imbalance in the degree to which graduate students in America are exposed to literature covering parts of the world beyond the West through reading lists and introductory course work.

There are several possible explanations for the scarcity of works outside the West on reading lists and syllabi. Works classified as "cross-regional" often pay more attention to developing regions, something that is not necessarily reflected in the results reported. Comparativists may be inclined to do work about countries in the West because of incentives for them to learn languages that can be used across a wider range of countries. Speaking Spanish probably offers greater research latitude than Farsi, Pashtu or Hangul. (This does not, however, explain the lack of attention to the Arabic-speaking world or the great attention to Western Europe). Once scholars have acquired a regional specialty, they may be disinclined to invest their time and effort in thoroughly understanding other parts of the world. Alternatively, scholars of comparative politics may perceive that inherently more interesting or significant theoretical questions have been addressed by works covering Western countries, which are not evident in research on other parts of the world. This leads them to assign more works covering Western industrial democracies than other parts of the world. Finally, it may be that Europeanists are disproportionately represented in the departments in our sample. These explanations, however, are *post hoc* and are best located somewhere on a continuum between tentative hypotheses and mere speculation.

Regardless of whether the foregoing explanations have any traction, the bias in the assignment of regional works is potentially problematic. Since introductory courses and the process of reading for comprehensive exams can contribute to the way in which dissertation topics emerge and embryonic research programs are conceived, this reality could have the effect of perpetuating further research on regions that have already received much attention, to the detriment of understudied areas. This may also be linked to job market opportunities, as more jobs may be available for Western Europe specialists, further enhancing the incentives for doctoral students to study already often-researched regions. This bias is even more troubling if one assumes that American graduate education in comparative politics should contribute to building intellectual capacity for addressing problems in states around the globe. Of course, there are avenues beyond comprehensive exams and survey courses that can spark graduate interest in regions of the world outside the industrialized Western democracies, but the training of regional specialists is certainly not commonly advanced by the aspects of graduate education discussed here.

10. Conclusion

Our findings suggest that there are areas of both agreement and disagreement in the way the field views the initial graduate training of comparativists. On the one hand, there appear to be very few works that are widely assigned across departments and considered essential reading in preparation for comprehensive exams. On the other hand, the work of some authors commands wide authority, and there appears to be relatively broad agreement about the themes (Political Economy, Regimes and Democratization, Methodology) as well as the regions (Western Europe, Latin America) emphasized. Nevertheless, we recognize that training in comparative politics involves much more than introductory coursework and preparation for comprehensive exams. The process of graduate education in comparative politics also involves taking advanced coursework, selecting a dissertation topic, doing field research, training in methods and languages, and the like. These factors both broaden and deepen a comparativist's training and allow him or her to find an area or theme of specialization. In light of this reality, students may have great latitude in fashioning their intellectual development, but introductory courses and exam preparation can have important implications for the formation of comparative politics scholars. Consequently, while the imbalances we have found in the early training of comparativists may be corrected by the entire process of graduate education, scholars may need to consider whether these imbalances should even exist.

Notes

[1] We thank all the departments and individuals whose participation made this research possible. We received reading lists from Michigan, Santa Barbara, Syracuse, Illinois, Florida, Notre Dame, Rochester, Texas, Washington, Yale, Harvard, Virginia, Princeton, Florida State, Minnesota, North Carolina and Northwestern; and syllabi from Arizona, Berkeley, Columbia, Duke, Emory, Georgetown, Indiana, Johns Hopkins, Maryland, Ohio State, Pennsylvania, Pittsburgh, UCLA, Washington University and Wisconsin-Madison. Of the 20 departments not included in our survey, 13 replied that they do have not have a reading list nor did they provide a syllabus from a comparative survey course, 6 did not respond and 1 (Tufts) had no Ph.D. program. These data were compiled into a large spreadsheet, and are available (without identifying particular departments) on the web at www.nd.edu/~apsacp.

[2] In a small number of instances, unpublished manuscripts appeared on the lists we gathered.

[3] Given that more than one thousand readings were collected, we turned to the comparative politics faculty in the Department of Political Science at the University of Notre Dame to assist in coding some readings by theme and region when we could not do this because of our unfamiliarity with the piece or the ambiguity in the title. Overall, eighty-four percent of the works were assigned a thematic code, while forty percent were given a regional designation, possibly because a fair number did not have a substantive regional focus.

[4] The Spearman correlation between the ranks of works on the reading lists and the syllabi is .444 for the top 30 pooled works and .650 for the top 20.

[5] Percentages do not add to 100, because some works were not classified by topic and because there is some overlap in many categories, since up to three topic codes could be assigned to a single work.

[6] Canonical works are assigned by more than one-third of the departments in the sample. Works assigned by half of the respondents are denoted with an asterisk (*), while those assigned by more than three-fourths of the respondents are denoted with two asterisks (**).

New Dataset Section

Michael Coppedge
University of Notre Dame
coppedge.1@nd.edu

Beginning with this issue, the datasets section will be a regular feature of *APSA-CP*. In this section we will publish links to archives of datasets, descriptions of recently archived comparative datasets and reviews of potentially important datasets. We welcome contributions from all of our readers, both faculty and graduate students. Any dataset cited in this section will be linked to the *APSA-CP* website at http://www.nd.edu/~apsacp/data.html.

Because dataset reviews are a new genre, some ground rules may be helpful. Like a book review, a dataset review should describe the contents in a way that helps potential users decide whether the data are relevant to their research. At a minimum, a review should describe the units (countries, groups, individuals), sample (e.g., how many countries or which regions are included or omitted), and time period. It should also give a fairly complete summary of the kinds of variables contained in the dataset. Ideally, a review should pay close attention to any innovative indicators in the dataset: indicators of concepts that had not been measured, or measured well, previously. It would be most helpful if the reviewer were to contrast such indicators with similar indicators available in other datasets. The reviewer should evaluate the validity and reliability of the key indicators. Do they measure what they claim to measure? Does the operational procedure inspire confidence? Were any reliability checks performed? Do the codings have face validity? In addition, reviewers must provide a web address for accessing the data and assess the ease of downloading the dataset. Reviewers may also want to provide information about the researchers who gathered the data and publications that have used the dataset. Finally, as with book reviews, dataset creators will not be allowed to review their own creations. Dataset creators may, however, submit brief (150-300 words) announcements describing their datasets and informing readers how to access them. The editors will publish announcements that are of general interest, space permitting.

To inaugurate this space, below are brief descriptions of a few of the larger dataset archives that are of special interest to comparativists.

1. Jennifer Widner has compiled a page with approximately 80 links to important or unusual datasets useful to comparativists. Send ideas to Professor Jennifer Widner (jwidner@umich.edu). http://www-personal.umich.edu/~jwidner/comparative/data.htm.

2. Richard Tucker, an international relations specialist at Vanderbilt, has compiled a "Replication Data Sets Archive" that offers links to many of the datasets used in publications. Many of these are useful for research in comparative politics or international relations.
http://www.vanderbilt.edu/~rtucker/methods/replication/

3. The Council of European Social Science Data Archives (CESSDA) "promotes the acquisition, archiving and distribution of electronic data for social science teaching and research." Its webpage links to 14 ICPSR-like data archives (including ICPSR) in the U.S. and Canada, 21 in Europe, and 5 others in Australia, New Zealand, South Africa, Israel, and Uruguay.
http://www.nsd.uib.no/cessda/europe.html

4. The UCSD Social Science Data Center's "Social Science Data on the Internet" page allows one to search or browse 748 Internet sites of numeric Social Science statistical data, data catalogs, data libraries, social science gateways, addresses, and more.
http://odwin.ucsd.edu/idata/

In this Newsletter:

How to Subscribe

Subscription to the APSA-CP Newsletter is a benefit to members of the Organized Section in Comparative Politics of the American Political Science Association. To join the section, check the appropriate box when joining the APSA or renewing your Association membership. Section dues are currently $7 annually, with a $2 surcharge for foreign addresses. The printing and mailing of the Newsletter is paid for out of member dues. To join APSA, contact:

American Political Science Association
1527 New Hampshire Ave., NW
Washington, DC 20036, USA

Telephone: (202) 483-2512
Facsimile: (202) 483-2657
e.mail: membership@apsa.com

Changes of address for the Newsletter take place automatically when members change their address with the APSA. Please do not send change-of-address information to the Newsletter.

APSA-CP Newsletter
University of Notre Dame
Decio Hall, Box "D"
Notre Dame, IN 46556
USA

University of Notre Dame
Non Profit Organization
U.S. Postage
Paid