

APSA-CP

Newsletter of the APSA Organized Section in Comparative Politics

Volume 9, Number 1

Winter 1998

Contents

News and Notes	3
Discussion	
<i>Comparative Method in the 1990s</i>	
David Collier, <i>introduction</i>	1
Andrew Bennett and Alexander George	6
Richard Locke and Kathleen Thelen	9
Tim McKeown	12
Roger Peterson and John Bowen	15
Charles C. Ragin	18
John D. Stephens	22
D. Stephen Voss and David Lublin	25
Datasets & Archives	
David Laitin	31
Good Reads	
David Laitin	33
Book Reviews	
<i>Traditional Politics and Regime Change in Brazil</i>	34

How to Subscribe

Subscriptions to the *APSA-CP Newsletter* are a benefit to members of the Organized Section in Comparative Politics of the American Political Science Association. To join the APSA, contact:

American Political
Science Association
1527 New Hampshire Ave., NW
Washington, DC 20036
Telephone: (202) 483-2512
Facsimile: (202) 483-2657
Email: membership@apsa.com

Letter from the President

Comparative Method in the 1990s

David Collier

University of California, Berkeley – dcollier@socs.berkeley.edu

The late 1960s and early 1970s saw the emergence of a new literature that established the “comparative method” as a fundamental component of the comparative politics enterprise. Comparative method was viewed as the systematic analysis of a relatively small number of cases (i.e., a “small *n*”), and was understood in contrast to the statistical, experimental, and case-study methods.

A quarter of a century later, we are now in the midst of a major new round of debates on this branch of methodology, and I wish to use my first letter from the president to make some observations about these debates. I focus here on what may be thought of as the division of labor in comparative politics between the comparative method and the statistical method, and also on the issue of conceptual validity, a long-standing concern of the comparative method. I will refer in my discussion to six articles in this issue of the *Newsletter* that reflect important facets of these debates.

Comparative Method vis-à-vis Statistical Method

How should we understand the role of the comparative method in relation to the statistical method? One view was offered in Arend Lijphart’s seminal article on “Comparative Politics and the Comparative Method” (*APSR*, 1971). Lijphart in effect saw the comparative method as a way station, at which analysts may stop to carry out initial tests of important hypotheses. Later, after scholars have done the hard work to create more sophisticated data sets, they should move on to research designs based on stronger empirical tests, utilizing the statistical method. According to this initial formulation of Lijphart’s view, the comparative method should play an important, but perhaps transitional, role within any given substantive area of research.

Given that many scholars believe that the statistical method is “obviously” a stronger approach, it is important to emphasize that Lijphart subsequently called attention to strengths and weaknesses of both the comparative and the statistical method. He underscored, among other things, the advantages of the comparative method in dealing with problems of conceptual validity, suggesting that perhaps we need to think of the comparative method as more than just a way station.

In that spirit, I view the comparative method as an important approach in its own right, one that is not limited to transitional or exploratory work. Within the field

Copyright 1998 American Political Science Association.

Published with the financial support of the
College of Letters and Science, UCLA.

Section Officers

President

David Collier

University of California, Berkeley
dcollier@socs.berkeley.edu

Vice-President and President-Elect

Michael Wallerstein

Northwestern University
mwall@merle.acns.nwu.edu

Secretary-Treasurer

James Caporaso

University of Washington
caporaso@u.washington.edu

1998 APSA Program Coordinator

George Tsebelis

University of California,
Los Angeles
tsebelis@ucla.edu

At-Large Committee Members

Barry Ames

University of Pittsburgh
barrya@pitt.edu

Nancy Bermeo

Princeton University
bermeo@princeton.edu

Ronald Herring

Cornell University
rjh5@cornell.edu

Susan Shirk

University of California,
San Diego
sshirk@ucsd.edu

Sidney Tarrow

Cornell University
sgt2@cornell.edu

of comparative politics, it remains a central methodology which scholars employ to accomplish important analytic tasks, and to which they periodically return, even at more “advanced” stages of research.

The cycle of returning to the comparative method takes various forms. First, it can be seen in the evolution of research on specific substantive topics. In a given area of study, a phase of research based on statistical analysis may be *followed*, rather than *preceded*, by a phase in which small-*n* comparison adds crucial insights. Scholars routinely go back to a small number of cases to assess the validity of conceptualization and measurement, as well as to refine causal inferences. Thus, small-*n* analysis has an important role to play, even when data for large-*n* studies are available.

A recent example of this sequence is found in the democratic peace literature, which analyzes the apparent tendency of democratic countries to go to war less frequently, at least with one another. The Bennett and George article below argues that an initial phase in this literature based on statistical analysis has been complemented by subsequent work in the comparative case-study tradition. Another example is found in the literature on the political economy of advanced industrial societies, in which a central goal has been to evaluate political explanations of national economic performance. In these studies, following an expansion of the *n* and a shift to more complex statistical modeling based on pooled time-series cross-section data, concern has subsequently been expressed about the reliability of causal inferences drawn from this type of data. One possible route to follow in light of

this concern is a new iteration of small-*n* research.

The recurring importance of the comparative method is also evident in the trajectory of methodological discussions. In debates of the 1990s on the relationship between quantitative and qualitative research, scholars have repeatedly gone back to insights drawn from the comparative method. The contributions below by Charles Ragin, John Stephens, and Timothy McKeown reflect these debates. Ragin compares the approach to causal assessment adopted by the comparative method with that of the statistical method. He highlights the problem of establishing “sufficient” causes and argues that this type of causation is more effectively analyzed by a new approach to the comparative method – based on “fuzzy logic” – than by statistical analysis. Stephens shows how the comparative method and the statistical method deal with the small-*n* problem, Galton’s problem, and the “black box” problem, offering the interesting observation that these two methods can suffer from similar dilemmas of indeterminacy in causal inference. McKeown adopts a different point of departure within the spectrum of methodologies, focusing on how causal inferences can be constructed on the basis of evidence and hypotheses derived from a single case. He contrasts this case-based approach with the statistical approach to causal inference, and his contribution serves as a useful reminder of the degree to which comparative work ultimately rests on the meticulous interpretation of individual cases.

A return to the comparative method is likewise seen in the trajectory some

(continued on page 4)

Use the Newsletter in the classroom. The APSA has authorized university teachers to reproduce articles from the Newsletter for use in the classroom at no charge. Take advantage of this policy, and introduce your graduate students to the latest research, issues and debates in comparative politics.

The Newsletter is available on the Internet for viewing and downloading.

Visit us at <http://shelley.polisci.ucla.edu/apsacp/index.html>.

Please contact the Assistant Editor, David Yamanishi, at falstaff@ucla.edu for assistance.

News & Notes

Newsletter Staff:

Editor

Miriam Golden
University of California, Los Angeles
golden@ucla.edu

Assistant Editor

David Yamanishi
University of California, Los Angeles
falstaff@ucla.edu

Regional Editors-at-Large:

Soviet Successor States

Daniel Treisman
University of California, Los Angeles

Middle East

Leonard Binder
University of California, Los Angeles

Latin America

Barbara Geddes
University of California, Los Angeles

Africa

Edmond Keller
University of California, Los Angeles

Western Europe

Ron Rogowski
University of California, Los Angeles

Eastern Europe

Ivan Szelenyi
University of California, Los Angeles

Formal Analysis and Methodology

John Londregan
University of California, Los Angeles

This year's section awards committees are as follows:

Luebbert Book Award Committee

Mark Beissinger (Chair)
University of Wisconsin
mbeissin@facstaff.wisc.edu
Robert Jackman
University of California, Davis
rwjackman@ucdavis.edu
Jennifer Widner
University of Michigan
jwidner@umich.edu

Luebbert Article Award Committee

Scott Mainwaring (Chair)
University of Notre Dame
scott.p.mainwaring.1@nd.edu
Timothy Frye
Ohio State University
tim.frye@polisci.sbs.ohio-state.edu
Thomas Callaghy
University of Pennsylvania
callaghy@wwic.si.edu

Sage Paper Award Committee

Barbara Geddes (Chair)
University of California,
Los Angeles
geddes@ucla.edu
Michael Bratton
Michigan State University
michael.bratton@ssc.msu.edu
Kathleen Thelen
Northwestern University
thelen@merle.acns.nwu.edu

George Tsebelis of the University of California, Los Angeles, will organize the section's panels at the 1998 APSA Annual Meeting.

The Walker Institute of International Studies of the University of South Carolina announces a new Working Paper Series, "Global Perspectives on Regime Change, Transitional Cultures, and Social Movements." The series is devoted to exploring the causes and consequences of the increasingly international currents shaping the politics and cultures of nation-states. The Walker Institute is particularly interested in showcasing the current work of a diverse group of scholars from the disciplines of political science, anthropology, history, economics and sociology. Papers published in this Series address questions such as: problems of democratization, demographic transitions, ethnic conflict, and the dynamics of social movements. For an up-to-date listing of Working Papers and ordering information, visit the Institute's web site: www.cla.sc.edu/iis/index.html. For further information, contact Dr. Maryjane Osa, Department of Government and International Studies, University of South Carolina, Columbia SC 29208 or email WIWPS@garnet.cla.sc.edu.

The Center for Development Studies is sponsoring two programs in Cuba. (1) In conjunction with the Facultad Latinoamericana de Ciencias Sociales, a travel and research seminar from July 5 to July 28, 1998 for professors and graduate students in the social sciences and history. (2) In conjunction with Presbyterian College, a six credit undergraduate course, including two weeks at Presbyterian College from May 17 to May 29 and four weeks in Cuba from May 31 to June 27, 1998. Both programs will be conducted in English. For more information concerning either program, contact Dr. Charles McKelvey, Center for Development Studies, 210 Belmont Stakes, Clinton, South Carolina, 29325; phone: (864) 833-8385 or (864) 833-1018; FAX 864-833-8481; e-mail: cemck@cs1.presby.edu.

**Please file changes of address
with the APSA.**

**Your Newsletter address will
change automatically.**

(Collier, continued from page 2)

times followed by specific research projects undertaken by individual scholars. Within a given study, a scholar focused on a small number of cases – for example, a limited number of national political regimes – may supplement the small-*n* comparative analysis of national units with further analysis focused *within* each country, based on a large *n*. Such within-case assessment might involve, for instance, analysis of public opinion data, national budgets, or other kinds of within-nation data that entail a large number of observations. However, to the extent that the goal is to bring explanatory insights from the within-case analysis back up to the level of the national political regimes that were the initial focus of concern, this ultimately remains a small-*N* analysis. Hence, the scholar will return to the comparative method in the final stage of the study.

Finally, the recurring importance of the comparative method is evident not only among scholars pursuing alternative methodologies, but also among analysts using diverse theoretical tools. For example, the forthcoming book summarized below by Peterson and Bowen includes five chapters in which game theorists test their models using carefully executed small-*n* comparisons.

To summarize, one sees not only periodic movement away from the comparative method, but also periodic movement back to it. Let me explore this theme further with reference to the issue of conceptual validity.

Conceptual Validity

Conceptual validity is an abiding issue in comparative research. The concern with validity is animated in part by a recognition of the trade-off between 1) the drive to extend our theories and hypotheses to a larger number of cases, and 2) the problem that if we extend them too far, conceptual stretching may occur, in that our concepts no longer validly fit our observations. This concern likewise derives from a fundamental preoccupation of many small-*n* analysts:

they worry that indicators employed in large-*n* cross-national research frequently fail to measure the concepts they purport to measure. Whatever vision one may have of the “scientific” status of comparative politics, this vision must include a central concern with validity. A focus on conceptual validity, correspondingly, has a prominent place in writing on comparative method. Major statements in the 1970s include Sartori’s analysis of conceptual stretching in “Concept Misformation in Comparative Politics” (*APSR*, 1970), and Przeworski and Teune’s recommendations in *The Logic of Comparative Social Inquiry* (Wiley, 1970) for adapting measurement to specific contexts, including potentially the use of what they call system-specific indicators.

Recent work has refined these perspectives in several ways. Charles Ragin has developed an analysis that parallels Sartori’s discussion of the intension (meaning) and extension (domain of relevant cases) of concepts. Ragin introduces the label “double fitting” to characterize the process of mutual adjustment between these two dimensions that often occurs in the course of concept formation. Shifts in meaning (i.e., in the definition of the concept) can push the analyst to adjust the corresponding domain of cases, and shifts in the domain of cases can necessitate an adjustment in the meaning, so as to maintain conceptual validity. Ragin suggests that in much research, as this double fitting proceeds, the domain of cases under investigation may remain fluid during initial phases of a study. Thus, in a comparative study of revolution, shifts in the definition of the main concept can dramatically change the relevant domain of positive and negative cases. Such shifts likewise occur in the broader evolution of scholarly research programs.

Given that establishing the domain of relevant cases is an essential underpinning for addressing various methodological issues, it is productive to recognize that this initial fluidity in defining this domain does indeed occur in many studies. It is impossible, for example, to make

judgements about selection bias until the domain of cases is established. A warning about another kind of bias is also essential. This process of double fitting should be used *appropriately* to refine concepts, and not *inappropriately* to come up with a set of cases that conveniently confirms the researcher’s preferred hypothesis.

A further contribution by Ragin to the discussion of validity is summarized in his article below. In a notable departure from his earlier focus on the dichotomous variables employed in Boolean algebra, he explores the possibility that the logic of fuzzy sets may sometimes offer a more valid operationalization of our concepts than does either dichotomous or quantitative measurement.

Another aspect of validity, linked to the idea of system-specific indicators, is explored below by Locke and Thelen. Whereas system-specific indicators were originally proposed as an approach to *quantitative* comparison, these authors suggest that scholars conducting *qualitative* research at times must engage in a parallel process of “contextualized comparison.” Thus, to generate conceptually equivalent observations in relation to a given concept, it is sometimes necessary to focus on what at a concrete level might be seen as distinct types of phenomena. For example, scholars who study national responses to external pressure for economic decentralization and flexibilization are sometimes concerned with identifying analytically equivalent “sticking points” where sharp conflicts emerge over this economic transformation. In the domain of labor politics such conflicts may, in different countries, arise over wage equity, hours of employment, work-force reduction, or shop-floor reorganization. The scholar must look at these different domains to make analytically equivalent comparisons that correspond to the concept of “sticking point.” Similarly, in *Shaping the Political Arena* (Princeton, 1991), Ruth Berins Collier and I applied the concept of the “initial incorporation” of the labor

movement in a parallel manner, recognizing that analytically equivalent observations linked to this concept entailed, in concrete terms, somewhat distinct phenomena in different countries.

Given the prominence of Przeworski and Teune's proposal for system-specific indicators, it is curious that in the intervening years this approach has not been used more frequently. Locke and Thelen's examples of comparing "privatization" and "globalization" across the countries of Eastern Europe help to clarify this puzzle. These examples suggest that comparativists who are closely familiar with the contexts they are comparing may in fact routinely employ this approach of contextualized comparison. Yet they often do so instinctively, rather than self-consciously. Following the phrase of Molière, it could be said that comparativists are sometimes "speaking prose" without recognizing it – i.e., carrying out contextualized comparison without being explicit about it. Clearly, it is preferable to make this practice explicit, and the Locke and Thelen article should help push scholars to do so.

Effective use of double fitting and contextualized comparison requires careful attention to the structure of concepts, to how concepts embody meaning, and to how scholars can most effectively use concepts in pursuit of their analytic goals. The recent small-*n* and case study literature on democratization offers examples of both successes and failures in the use of concepts. These successes and failures arise in part out of scholars' responses to two conceptual challenges posed by the recent world-wide wave of democratization. Analysts seek both to increase analytic differentiation in order to capture the diverse forms of democracy that have emerged, and also to avoid the conceptual stretching which arises when the concept of democracy is applied to cases for which, by relevant scholarly standards, it is not fully appropriate. A dilemma arises from the fact that efforts to increase differentiation through introducing finer distinctions may produce analytic categories that are more vul-

nerable to conceptual stretching.

Analysts have fine tuned their concepts in many different ways as they pursue these contending objectives, including the creation of what may be called "diminished" subtypes of democracy. For example, the concept of "illiberal democracy" can serve to differentiate cases where the protection of civil liberties is seen as inadequate; and because it is a diminished subtype, it avoids conceptual stretching by specifically *not* making the claim that these are full instances of democracy, which by standard definitions they clearly are not. In the hands of careful, well-disciplined scholars, such conceptual innovations can yield better research.

However, this proliferation of conceptual forms also has a down side. For example, the literature on democratization has spun out literally hundreds of democratic subtypes, and too often these subtypes either are not clearly defined, or are not employed in a consistent manner, or both. Consequently, any gains that might be achieved in finer analytic differentiation and/or improved conceptual validity may be cancelled out by the resulting conceptual confusion. When such confusion arises, it is essential for scholars to engage in a self-conscious, critical evaluation that systematically appraises existing usage of concepts and seeks to channel it in more productive directions.

Researchers who work closely with a small *n* are supposed to have the advantage of "knowing their cases," thereby helping them to avoid the problems of validity that may arise for scholars who are not as familiar with the contexts they are studying. Yet in addition to knowing their cases, scholars need a disciplined understanding of how to employ concepts, along with a firm grasp of how to organize concepts into worthwhile theoretical arguments. The challenges of learning and teaching these skills, as well as applying them effectively in different substantive domains of research, must be an abiding concern in the field of comparative method.

References

- Nathaniel Beck and Jonathan N. Katz, "What To Do (and Not To Do) with Time-Series Cross-Section Data," *American Political Science Review* 89, No. 3 (September 1995): 634-647.
- David Collier and Steven Levitsky, "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics* 49, No. 3 (April 1997): 430-451.
- Bernhard Kittel, "Sense and Sensitiveness in Pooled Analysis of Political Data," *European Journal of Political Research*, forthcoming, 1998.
- Arend Lijphart, "The Comparable-Cases Strategy in Comparative Research," *Comparative Political Studies* 8, No. 2 (July 1975): 169-171.
- Charles C. Ragin, *Constructing Social Research: The Unity and Diversity of Method* (Thousand Oaks: Pine Forge Press, 1994), chap. 4.

Book Reviewers Welcome

Doctoral students at any institution are welcome to submit book reviews. The review itself must be received by the editor before a decision to publish will be made. Books must have been published within the past two years. Reviewers are responsible for procuring their own copies of books. Reviews of foreign language material that might otherwise remain little known in the U.S. are encouraged. For more information, contact the editor or assistant editor.

An Alliance of Statistical and Case Study Methods: Research on the Interdemocratic Peace

Andrew Bennett

Georgetown University

benneta@gunet.georgetown.edu

Alexander George

Stanford University

Introduction

Claims that democracies are more peaceful than other kinds of regimes, or at least more peaceful toward other democracies, have recently engendered an active research program. This program illustrates particularly well the strengths and weaknesses of contemporary statistical and case study methods. Statistical methods dominated the first wave of research on the democratic peace. These methods have comparative advantages in identifying correlations among variables, controlling for the effects of rival hypotheses, and testing for possible spuriousness. In short, statistical methods are good at establishing that variables have measurable “causal effects,” or that changes in these variables are systematically related to changes in outcomes (King, Keohane and Verba, 1994: 76-82).

Case study methods, more prominent in the second wave of democratic peace research, are relatively weak at measuring causal effects. However, they have comparative advantages in areas where statistical methods are less effective. These include specifying and measuring complex qualitative variables, inductively identifying new variables and hypotheses, and developing contingent generalizations or typological theories. More generally, case study methods are strong at identifying and testing “causal mechanisms,” or the social or political processes through which variables exert causal effects (Yee, 1996: 69-85).

Yet as “scientific realists” have argued, explanatory theories require assertions about both causal effects and causal mechanisms. Statistical and case study methods are complementary in establish-

ing the different claims necessary for causal explanation. The democratic peace research program has thus progressed much farther through the combination of both methods than it would have through the application of either set of methods alone.

I) *The First Wave of Research on the Democratic Peace: Contributions of Statistical Methods*

Most of the first wave of research on the democratic peace up through the 1980s used statistical methods. Studies using these methods made three important contributions. First, they refined the research question, shifting the focus from the question of whether democratic states are more peaceful in general (the “democratic peace”), to whether they are more peaceful only or primarily vis-a-vis one another (the “inter-democratic peace”). An additional refinement concerned whether democracies are only less likely to fight wars with one another, or also less likely to engage in conflicts short of war. Researchers also began to examine whether sub-types of states, such as states in transition to democracy, were more or less war-prone.

Second, many statistical studies tested for whether findings of an inter-democratic peace were spurious. They did so by controlling for variables such as contiguity, wealth, alliance membership, relative military capabilities, rates of economic growth, and the presence of a hegemon.

Third, researchers using formal models as well as statistical methods deductively theorized about and empirically tested the potential causal mechanisms behind an (inter)democratic peace, often grouping them together under explanations relating to democratic norms and/or institutions (Maoz and Russett, 1993, illustrates all three of these contributions).

Statistical methods achieved important advances on the issue of whether a non-spurious (inter)democratic peace exists. A fairly strong though not unanimous consensus emerged that: (1) democracies are not less war-prone in gen-

eral; (2) they have very rarely if ever fought one another; (3) this pattern of an inter-democratic peace applies to both war and conflicts short of war; (4) states in transition to democracy are more war-prone than established democracies; and, (5) these correlations were not spuriously brought about by the most obvious alternative explanations.

However, statistical studies have proved more capable of addressing *whether* a non-spurious democratic peace exists than of answering *why* it might exist. Researchers deductively derived several potential causal mechanisms that might explain an (inter)democratic peace. Yet these mechanisms were often inconsistent, suggesting the possibility of different causal paths to an interdemocratic peace. Moreover, statistical methods proved inadequate to test these mechanisms for two reasons. First, statistical methods faced daunting problems in measuring variables like democratic norms and institutions. Second, statistical methods are not well-suited to identifying inductively or empirically testing causal mechanisms. These methods are optimal for assessing correlations, but correlations do not establish causality. The very techniques that make statistical methods powerful at testing for correlations across many cases make these methods unable to assess the causal mechanisms at work in a single case due to the lack of sufficient degrees of freedom for partial correlations.

In contrast, “case-oriented research” uses process tracing to test whether a proposed explanation is consistent with the evidence *in a given case*, and by extension whether the same causal process might apply to a category of cases with similar values on the independent variables. Process tracing is the method of looking at the observable variables along a hypothesized causal process through which a causal mechanism exerts an observed causal effect. It uses a logic that is fundamentally different from that of statistical correlations. If process tracing shows that a single step in the hypothesized causal chain in a

case is not as predicted, then an unmodified version of the hypothesis cannot explain that case, even if it does explain most or even all other cases. If there is only one intervening step in the hypothesized process, and this is observed to be untrue in the case, the hypothesis cannot explain that case. At the same time, if a complex causal hypothesis involves several steps and only one of these is observed to be inoperative, the hypothesis cannot explain the case.

In contrast, if a statistical test were (inappropriately) applied to such process tracing data, it would find insufficient degrees of freedom in the first instance, in which one variable did not fit, to reach any conclusions. In the second instance, where all but one of several intervening variables did fit, a statistical test might wrongly conclude that the process in question demonstrated a high and possibly causal correlation. The logic of testing causal mechanisms in particular cases, which requires the full consistency of all specified intervening variables, is thus quite different from that of establishing correlations on causal effects across many cases, which requires probabilistic associations.

A final factor makes the democratic peace research program amenable to case study methods. This is the fact that contiguous democracies and periods of war in a given dyad are rare relative to the large number of dyads in history. For statistical researchers, this is a limitation: given the small number of potential wars between democracies, the existence of even a few wars between democracies or the omission of a single variable could erase much of the statistical support for an interdemocratic peace. Because there are about twenty potential exceptions to the assertion that democracies have never fought wars with one another, the results of statistical studies must remain provisional despite the emerging consensus that an interdemocratic peace exists. (Ray 1995: 86-7) For case study researchers, this is an opportunity: it is possible for the field as a whole to study every one of the possible exceptions to the democratic peace, and to study com-

parative cases of mixed dyads and non-democratic dyads. Indeed, many of the possible exceptions to the democratic peace are already the subject of several case studies.

II) The Second Generation of the Research Program: Case Study Contributions

In the 1990s, the most pressing puzzle in the research program shifted from whether a democratic peace existed, a question for which statistical methods were well-suited, to why such a peace might exist, an issue best addressed through case study methods. One advantage of case studies is that they are better able to measure complex qualitative variables. For many of the dyads of interest to democratic peace researchers, polling data is not available, so in measuring democratic norms statistical researchers have had to use proxy variables such as the number of deaths or executions related to domestic violence. In contrast, case studies have drawn on more internally valid qualitative measures for democratic norms and institutions, such as the statements and writings of contemporary leaders and the detailed assessments of regional experts and historians.

A second advantage of case studies is their ability to identify additional variables inductively. Statistical methods can also identify new variables, but variables inductively identified through correlations alone may be spurious if they are not tied to causal mechanisms. Formal modeling can also identify new variables, but it relies primarily on deduction and requires subsequent empirical tests. In case studies, the inductive use of process tracing can turn up unanticipated variables that are directly tied to causal mechanisms. In the democratic peace research program, case studies have identified or tested several new variables, including issue-specific state structure, specific norms on reciprocity and the use of deadly force, leaders' perceptions of the "democraticness" of other states, transparency, and the distinction between status quo and challenger states. Nota-

bly, each of these calls forth a causal mechanism relevant to the interpretation of statistical correlations, rather than simply being an atheoretical induction.

Third, process tracing has proved a powerful method of testing claims about causal mechanisms related to the interdemocratic peace. There are still relatively few case studies on the democratic peace, and these studies have not yet established a consensus on which causal mechanisms might help account for an interdemocratic peace. Still, case studies have been able to rule out some causal mechanisms in important cases. For example, the assertion that democratic mass publics oppose wars with other democracies does not hold for the Fashoda Crisis, in which Britain and France avoided war despite the British public's support for using force.

Fourth, case studies can develop typological theories, in which different combinations of independent variables may interact to produce similar outcomes on the dependent variable. With many political phenomena, the same outcome can arise through different causal paths in which there may be no single non-trivial necessary or sufficient condition. This is known as "equifinality." The goal of case study researchers is thus not simply to affirm or reject the democratic peace as a valid correlation, but to identify the conditions under which specified types of democracies interact with systemic and other variables to produce specific types of conflict behavior in democratic or mixed dyads (Elman 1997: 6, 39-40). The resulting theories usually focus on interactions among combinations of variables, rather than variables considered alone or isolated through means of statistical control.

The development of typological theories thus involves the differentiation of independent and dependent variables into qualitatively different "types," such as types of war or types of democracy. The task of defining "war" and "democracy" is challenging for both statistical and case study researchers, and they respond to it differently. Statistical researchers attempt to develop rigorous but general

definitions, with a few attributes that apply across a wide number of cases. Case study researchers usually include a larger number of attributes to develop more numerous types and subtypes, each of which may apply to a relatively small number of cases (Collier and Levitsky 1997; Elman 1997: 35-40). In research on the democratic peace, case study researchers have suggested differentiating between centralized and decentralized democracies and among democracies where leaders and mass publics either converge or differ on norms regarding the use of force (Elman 1997:39-40). It may also prove useful to distinguish between “conditional peace” in a dyad, which depends on continued military deterrence, and the kind of “stable peace” in which the resort to force is not threatened. Researchers can then address whether joint democracy is necessary or sufficient for stable peace, and they can use process tracing to explore the conditions under which conditional peace can change into stable peace. For example, there may be no single combination of democratic norms and institutions that produces an interdemocratic peace, and the paths through which stable peace emerges in different dyads may vary depending on whether the development of democratic norms preceded or followed that of democratic institutions.

However, not every sub-type is useful. Researchers should not simply define away anomalies through the creation of sub-types. As a methodological safeguard, a new sub-type should not only survive statistical or process tracing tests, but should identify and then empirically verify hitherto unexpected observable implications. The assertion that “new” or “transitional” democracies are more war-prone, for example, posits testable correlations and causal mechanisms and suggests dynamics that should make states in transitions out of as well as into democracy more war-prone. The exclusion of civil wars from cases of democratic wars is more questionable, as is the exclusion of conflicts that fall below the arbitrary figure of 1,000 battle deaths

that is used in some data sets.

III) Critiques and Challenges of Case Study Methods as Applied to the Democratic Peace

Two dilemmas of case study methods are evident in the democratic peace literature: the problem of case selection and that of reconciling conflicting interpretations of the same cases. Researchers’ subjective biases may lead them to select cases that overconfirm their favorite hypotheses. This is a potentially more serious problem than that of selection bias in statistical studies, which tends to result in underconfirmation of hypotheses. Biased case selection can also arise from the fact that evidence on certain cases is more readily accessible than that on others, and from the tendency for historically important cases to be overrepresented relative to obscure but theoretically illuminating cases. For example, democratic peace case studies have overemphasized cases involving the United States.

On the positive side, there is an emerging consensus among supporters and critics of the democratic peace on which cases deserve study, demonstrating that case selection is not an arbitrary process. Several cases have been mentioned by numerous scholars as possible exceptions to the democratic peace, including the War of 1812, the American Civil war, conflicts between Ecuador and Peru, the Fashoda crisis, the Spanish-American War, Finland’s conflict with Britain in World War II, and a dozen or so other conflicts or near-conflicts (Ray 1995: 86-7). The initial focus on these “near wars” between democracies and “near democracies” that went to war was appropriate, as it offered tough tests of a democratic peace. As researchers accumulate adequate studies of these cases, they can branch out into more comparisons to mixed and non-democratic dyads.

A second challenge, that of judging conflicting interpretations of the same cases, arises from the fact that competing explanations may be equally consistent with the process-tracing evidence.

This makes it hard to determine whether both explanations are at work and the outcome is overdetermined, or whether the variables in competing explanations have a cumulative effect, or whether one variable is causal and the other spurious. Competing explanations may also disagree on the “facts” of a case or address incommensurate aspects of a case. Often, it is possible to reconcile differing interpretations by: (1) identifying and addressing factual errors, disagreements, and misunderstandings; (2) identifying all potentially relevant theoretical variables and hypotheses; (3) comparing various case studies of the same events that employ different theoretical perspectives; (4) identifying additional testable and observable implications of competing interpretations of a single case; and, (5) identifying the scope conditions for explanations of a case or category of cases (Njolstad 1990: 240-244).

Examples from the democratic peace literature illustrate how these suggestions work in practice. There is some factual disagreement on whether both British and French public opinion was bellicose in the Fashoda crisis, or whether British public opinion was substantially more supportive of going to war. Some argue that foreign policy-making was so dominated by elites in both cases that public opinion made little difference. Similarly, there is some disagreement on the nature and salience of public opinion in Spain at the time of the Spanish-American War. Additional historical research might help resolve these issues.

On the Fashoda crisis, there is disagreement on whether joint democracy and a wide power imbalance overdetermined the peaceful outcome, whether they had cumulative effects, or whether one factor was causal and the other spurious. More systematic analysis of process tracing data, or careful counter-factual analysis, might resolve this controversy, although it is also possible that no scholarly consensus will emerge. The same is true of discussions on whether a large power disparity and the (perceived) absence of democracy in Spain were jointly necessary for the

Spanish-American War. In case study methods, as in statistical methods, scholars may at times have to live with some degree of indeterminacy when competing variables push in the same direction.

Conclusions

We use the democratic peace research program as a methodological example not because its historical evolution is typical but because it illustrates particularly well the strengths and limits of both methods. Our argument does not imply that case study methods will supplant statistical studies in this program, or that the historical evolution of social science research programs is usually from quantitative to qualitative methods. Usually research using both methods proceeds simultaneously and iteratively, as each method confronts new research tasks at which the other method is superior. Indeed, as case study researchers devise more differentiated measures of “democracy,” their findings will no longer enjoy the empirical support of statistical methods using the definitions employed in existing databases. New statistical studies will need new databases using the refined definitions. Formal modeling can also help identify possible counterintuitive dynamics on the democratic peace that can be submitted to empirical testing by statistical and/or case study methods.

The evolution of this research program does not suggest that case study methods are somehow “better” than statistical methods, any more than the reverse. Rather, the two methods’ contributions are complementary but not identical. They provide epistemologically different types of knowledge. Statistical methods have more effectively addressed the question of whether a democratic or interdemocratic peace exists – corresponding to the notion of causal effects. Case study methods have been more effective at testing the proposed reasons for why such a peace might exist – corresponding to the notion of causal mechanisms. Adequate causal explanations must include assertions on causal effects *and* on the underlying causal

mechanisms that bring about these observed effects. Theological arguments that causal effects are “logically prior” to causal mechanisms (KKV, 1994: 76-82), or that causal mechanisms are “ontologically prior” to causal effects (Yee, 1996: 69-85), miss the point. Neither of these components of explanatory theory, and neither of the methods best-suited to capturing them, should be privileged over the other.

References

- David Collier and Steven Levitsky, “Democracy With Adjectives: Conceptual Innovation in Comparative Research,” *World Politics* Vol. 49, no. 3 (April, 1997).
- Miriam Elman, ed., *Paths to Peace: Is Democracy the Answer?*, (Cambridge, MA, MIT Press, 1997).
- Gary King, Robert Keohane and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ, Princeton University Press, 1994).
- Zeev Maoz and Bruce Russett, “Normative and Structural Causes of the Democratic Peace,” *American Political Science Review* Vol. 87, No. 3 (September, 1993).
- Olav Njolstad, “Learning From History? Case Studies and the Limits to Theory Building,” in Njolstad, ed., *Arms Races: Technological and Political Dynamics* (Sage, 1990).
- James Lee Ray, *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition* (Columbia: University of South Carolina Press, 1995).
- Albert Yee, “The Effects of Ideas on Policies,” *International Organization* Vol. 50, No. 1 (Winter, 1996).

Note: We wish to thank David Collier for his suggestions on this paper, a fuller version of which is available at <http://www.georgetown.edu/bennett>.

Problems of Equivalence in Comparative Politics: Apples and Oranges, Again

Richard Locke

Massachusetts Institute of Technology
rlocke@mit.edu

Kathleen Thelen

Northwestern University
thelen@merle.acns.nwu.edu

Introduction

The past several years have witnessed lively debates in comparative methodology focusing on important issues such as case selection and the relative strengths of qualitative versus quantitative research strategies. This research note takes up an issue that recent methodological debates have largely skirted or ignored, namely the question of issue or process equivalence in cross-national comparative research. How to compare “like with like” is a very old problem in comparative research. In their classic *The Logic of Comparative Social Inquiry*, Adam Przeworski and Henry Teune discuss at length the problem of establishing equivalent cross-national indicators and measures. We believe that their admonitions have been largely unheeded in a good deal of comparative research, which has been insufficiently concerned with this problem and altogether too quick to assign equivalence to processes whose meaning may well vary when situated within different contexts.

Our argument is two-fold. First, we suggest that comparative research needs to attend more closely to the question of whether “matched comparisons” that track the same phenomenon or process in different contexts are in fact comparing apples with apples. Second, we argue that in order to answer certain types of questions, a different research strategy may be required, one which compares “apples with oranges”, that is, looks at different processes in different countries, in order to capture analytically equivalent issues. In short, a more

“contextualized” approach to comparative research is required to both address the issue of equivalence and fully leverage the analytic power of qualitative comparisons. We develop this argument in three steps. First, we address the general issue of equivalence, drawing especially on Przeworski and Teune’s discussion. Second, we provide an example that illustrates the importance of “contextualized comparison.” Third, continuing this example, we suggest that “contextualizing” comparisons may in some cases involve a research strategy that looks at different processes rather than the same process cross-nationally.

The Problem of Equivalence in Comparative Politics.

Przeworski and Teune emphasize equivalence of indicators and of measurement in cross-national research, arguing that “problems of measurement arise in comparative research largely from the need to incorporate *contextual* characteristics of complex systems into the language of measurement” (p. 92, italics ours). Noting that “the cultural or societal contexts in which ... observations are made may distort the validity of the inference” (p. 94), they stress that “validity means that we are measuring in *each system* under consideration what we intended to measure” (p. 103, emphasis in original). The authors give the example of political participation; because politics are organized very differently in different countries, political participation is not “expressed in terms of the same behavior” in all countries, and thus in each, it needs to be measured in context-appropriate ways. Rather than assume that a phenomenon can be measured with a single, standardized measure or indicator, Przeworski and Teune clearly put the onus on the researcher to be sensitive to what they call “system interference” and to make adjustments if necessary to establish equivalence (103ff). As they emphasize, the central concern in formulating theories is generality, but when it comes to establishing valid measures and indicators analysts need to attend to the specific features

of a given system.

We wholeheartedly agree with Przeworski and Teune but believe that as common-sensical as this advice may sound, most comparative research has largely ignored it. For example, quantitative studies and survey research routinely use standardized indicators of complex social processes without considering whether or not they are really tapping the same process in different contexts. As Rueschemeyer and Stephens point out, “cross-national statistical research settles on one standardized operationalization and takes inadequacies of fit, which vary across cases, into the bargain.” They suggest that “qualitative comparative historical research can give much closer attention to the match between evidence and theoretical conceptualization.”¹ Unfortunately, this potential is not always exploited, and much contemporary qualitative work is equally quick to rely on “matched comparisons” that track a given phenomenon in different countries without considering how the same process or phenomenon can have contrasting meanings in different contexts.

To illustrate, let us take an example from the literature on contemporary labor politics in the advanced industrial democracies.² The dominant approach in this literature has been to fix on a single issue or process (e.g., wage bargaining or work reorganization) and to compare developments in the selected area across a range of countries. Thus, we have important quantitative studies of trends in wage bargaining across a large number of countries, as well as more qualitative studies that compare work reorganization in different national contexts. Such “matched comparisons” have taught us a great deal about the relative success and failure of unions in different countries to cope with particular changes. Yet because we do not know whether these issues have the same meaning or importance in each of the countries being compared, we have no idea whether or not the various unions were, in fact, fighting the same battle. In short, what many of these studies by

and large do not consider – and indeed, what the research design itself obscures – is that the very same issue may have a very different meaning or valence in different countries and hence, quite logically, provoke very different outcomes.

Take, for example, the issue of work reorganization. A large literature tells us that one of the most serious challenges facing unions in the advanced industrial countries is employer efforts to reorganize work along more “flexible” lines. Indeed, matched comparisons reveal broad differences in the ability or success of unions in different countries to cope with this common trend. In countries such as Sweden and Germany, studies show that unions have been active participants in workplace restructuring, whereas in the United Kingdom and the United States, the reorganization of work has often undermined union strength and thus prevented unions from influencing the content and direction of change on the shop floor.

This is all very interesting and true. But before we draw any broad lessons from these divergent experiences, we need to consider explicitly the contrasting meaning or valence of work reorganization in these different countries. In fact, the significance of shopfloor reorganization varies tremendously from country to country. Unions in the United States have strongly resisted more flexible forms of work organization, because this kind of change undermines narrow job definitions with their related wage, seniority, and security provisions – practices that represent the institutional anchors for American unions’ traditional rights within the firm. In Germany, by contrast, where employment security and union strength are not dependent on shop-floor practices such as job control, works councils and their unions have welcomed similar changes that upgrade their skills and enhance their autonomy. This example illustrates how the very same issue or process can have distinct meanings in different national settings, depending on contrasts in institutional starting points and in the impact of various changes on traditional arrangements.

Where this is the case, the conventional practice of comparing apparently similar changes across countries and attributing varying degrees of labor “success” to different national institutional arrangements is somewhat misleading. These comparisons are misleading because they give the impression that they are comparing “apples with apples” when instead, given differences in starting points and varying degrees of valence different issues possess in different national contexts, they are often in practice comparing substantially different phenomena. By failing to confront the issue of equivalence, matched comparisons of this sort frequently blend out important differences in starting points that may in fact hold the key to explaining the observed divergent outcomes.

Contextualized Comparison

Rather than assume (or arbitrarily assign) equivalence to the same process cross-nationally, we need to ask specifically if we are in fact comparing like with like. And indeed, in order to answer certain kinds of questions, an entirely different approach to comparative analysis may be required. What we have called “contextualized comparison” is a strategy which self-consciously seeks to address the issue of equivalence by searching for *analytically equivalent* phenomena – even if expressed in substantively different terms – across different contexts. Analysts interested in the relative success of different union movements in dealing with common pressures for decentralization and flexibility need to be aware that these common international trends have been refracted into very different conflicts, centering on divergent substantive issues in alternative national contexts. National institutional arrangements create different sets of rigidities and flexibilities in different countries, so that conflicts between labor and management have come to center on different “sticking points.” Thus, if we want to know about how well unions are succeeding cross-nationally, it may be more appropriate to compare across these “sticking points” rather than to track a

single issue, like work organization, cross-nationally.

To illustrate the point, let us return briefly to the previous example. Work reorganization has been a much more conflictual “sticking point” between labor and employers in the United States than in many other countries. Work rules and job classifications gave organized labor in the U.S. a set of rights and an established role (monitoring these rules) within the firm. As a result, work reorganization aimed at eliminating these rules and classifications threatens to alter if not eliminate established union rights and hence union presence on the shop floor. This is why these issues have so much more valence and have provoked so much more conflict in the United States than other, analogous changes in other American industrial relations practices (e.g., hiring and firing practices, flexible compensation schemes, contingent employment arrangements) and also why this same issue is less contested elsewhere. In Germany, for instance, work reorganization is not tied up with a similar reordering of core union rights; in fact (in stark contrast to the United States) German unions had reasons of their own for embracing and actively promoting work reorganization.

At the same time, however, *other* issues have indeed been important sticking points between unions and employers in other countries. Wage flexibility, for example, was a hotly contested issue in Sweden in the 1980s because of traditional bargaining structures and union policies that were both premised on and sustained a much higher degree of wage compression than in the United States. In Germany, the relative ease with which work reorganization has been negotiated contrasts sharply with the major struggles between German unions and employer associations over shorter and more flexible work hours.

In short, putatively common international trends have in fact set in motion rather different conflicts in different national contexts. Conventional studies often draw broad conclusions concerning relative union success or failure from an

analysis of a single issue area in different countries. But the choice of which issue area to study can affect the results considerably. As we have seen, German unions have been more successful in negotiating changes in work reorganization, but then again, work reorganization does not pose the same kinds of problems for German unions that it does for American unions. The relative ease with which work reorganization in Sweden has been negotiated contrasts sharply with the considerable conflict other issue areas, such as wage flexibility, have sparked in that country. As a result, focusing on work reorganization alone tells us little about how well Swedish or German unions do when employers’ goals clash more directly with the traditional institutional foundations of union power. The strategy of contextualized comparison confronts this issue by explicitly considering cross-national variation in conflicts centering on (different, nationally specific) sticking points.

The strategy of contextualized comparison is not limited to labor scholarship but has broader implications for other specializations in comparative politics. Consider, for example, the debates surrounding the economic and political reconfiguration of Eastern Europe. Some of this work has sought to assess the relative “success” or “failure” of the transformation process in different East European countries by analyzing how far along they have come or well they are doing, for instance, in promoting privatization or in achieving macroeconomic stabilization.³ Differences along a supposedly common trajectory are often seen to indicate varying degrees of political will or commitment to democratic capitalism. But this assumes that these countries embarked on these various processes from the same point of departure, and this is clearly not the case. Other work – intuitively if generally not explicitly guided by some of the points we are stressing here – recognizes that “privatization” or “macroeconomic stabilization” not only involve different kinds of policy initiatives, but also have very

different valence in different national contexts, depending among other things on the policy legacies of their previous regimes. For example, Hungarian communism had long tolerated and even promoted a vibrant second (private) economy and, thus, privatization was less of a sticking point and also takes a very different form than in, say, Bulgaria or even Eastern Germany, where privatizers had no strong indigenous model to build on. Likewise, macroeconomic stabilization posed significantly greater challenges for Poland, given its enormous foreign debt, than it did for the Czech Republic, with its long-standing tradition of fiscal conservatism.⁴ In short, to understand the politics surrounding the transformation process in Eastern Europe, scholars need to be sensitive to the contextual conditions that frame the putatively common challenges that these various economies face, and also explicitly recognize the varied valence that particular issues have in different countries.

Consider another example, the “globalization” of different national political economies. Again, many analyses of this phenomenon portray it as if the same set of external pressures (e.g., increased international competition and trade, financial interdependence, supranational regulatory shifts, etc.) are equally pervasive or intense for all national economies. But this is not so. Countries differ not only in their historical legacies and current institutional arrangements but also in their place within the international division of labor. The seemingly “common” challenge of globalization is in fact refracted into very different kinds of problems in different systems – depending, for example, on the degree and type of openness of various economies, the size of their domestic markets, what sectors are competing internationally, and on what terms these sectors are competing (cost-based versus product differentiation strategies). Any effort to assess the overall impact of globalization thus needs to be sensitive to variation in where its impact is being felt. As Robert Boyer and Wolfgang Streeck have recently

pointed out,⁵ in Germany, the problems posed by globalization present themselves – first and foremost – in labor market institutions, whereas in Japan, it is the financial system that has emerged as the “weak link” in the system. For anyone interested what globalization means for the advanced industrial democracies, this is already an important insight, one that our research should acknowledge and build upon rather than obscure.

In sum, contextualizing comparative analysis means not simply being more careful about our choice of categories or phenomena to compare, or about the importance of issue or process equivalence; but may also push us at times to make different kinds of comparisons altogether. What, at first, might look like “apples and oranges,” may turn out to be, under closer examination, a more effective way of capturing the particular way common challenges have been translated into specific conflicts in the various national settings. This more nuanced and context-sensitive approach to issues of equivalence, we believe, is among the greatest contributions that qualitative comparative analyses can make to our field.

¹ Dietrich Rueschemeyer and John D. Stephens, “Comparing Historical Sequences – A Powerful Tool for Causal Analysis,” *Comparative Social Research* 16 (1997), p. 58.

² For a more extended discussion of these comparisons of labor politics, see Richard M. Locke and Kathleen Thelen, “Apples and Oranges Revisited: Contextualized Comparisons and the Study of Comparative Labor Politics,” *Politics and Society* 23:3 (September 1995), 337-367.

³ For example, Oliver Blanchard, Kenneth A. Froot, and Jeffrey D. Sachs, *The Transition in Eastern Europe* (University of Chicago Press, 1994).

⁴ For more on how seemingly similar processes play themselves out quite differently in different east European countries, see Anna Seleny, “Old Political Rationalities and New Democ-

racies: Compromise and Confrontation in Hungary and Poland,” unpublished manuscript, Department of Politics, Princeton University, July 1997; and David Stark, “Heterarchy: Asset Ambiguity, Organizational Innovation, and the Postsocialist Firm,” unpublished manuscript, Department of Sociology, Columbia University, November 1997.

⁵ Comments at Workshop on Japan and Germany, Seattle, Washington (April 1997).

Why Is A Single Case Important?

Tim McKeown

University of North Carolina
tim_mckeown@unc.edu

Debates on the scientific status of rival methodologies have profound implications for the kind of social status, financial and institutional support customarily bestowed upon them. Disputes over the scientific status of case studies have unfortunately taken for granted the idea that the philosophical and methodological presumptions that are commonly taken to underlie statistical analyses of data also provide the proper foundation for the evaluation of case studies. Even defenders of case studies typically couch their defense in terms of the language of covering laws, falsification, degrees of freedom and the like.

It is well to consider whether such a point of view is really warranted. The pastiche of positivist and Popperian positions and classical statistical theory commonly invoked as the basis for evaluating case studies is itself problematic. I have addressed these problems elsewhere; here I wish to present a different way of thinking about case studies – one that seems to accord much more closely with practitioners’ self-understandings, as well as providing more helpful and less distorting guidance for the conduct of research than the statistical metaphor does. To claim that inferences are drawn and tested is not to claim that they are tested using a process that mimics classical statistics or relies only on

the results of statistical tests.

Stephen Toulmin has suggested that legal proceedings be taken as an exemplar of how a community arrives at judgements about the truthfulness of various statements. In such proceedings judges or juries are asked to make judgements about causation and intent based quite literally on a single case. Although statistical evidence sometimes is used in court, the only way that judicial decisions are statistical in any more general sense is in the implicitly probabilistic conception of guilt that underlies an evidentiary standard such as “beyond a reasonable doubt.” Likewise, if one considers the standard set of successful scientific research programs that are commonly used as exemplars in discussions of the philosophy of science, one searches in vain among these cases from early modern chemistry, astronomy or physics, from the germ theory of disease or the theory of evolution, for any instance where explicit statistical inference played a noticeable role in the development of these research programs. If all our understandings of the world are statistical, then it is difficult to see how any judge or jury could ever convict anyone (unless perhaps the defendant were being tried for multiple crimes). If there is a statistical logic to all scientific inference, what are we to make of situations in the physical or biological sciences where a few observations (or even a single one in the case of Einstein’s theory of relativity and the bending of light by gravity) in non-experimental situations were widely perceived to have large theoretical implications? While there are all sorts of criticisms that are leveled against judicial systems, I am aware of no one who claims that judges and juries are literally incapable of coming to defensible judgements about guilt or innocence on the basis of a single case. Likewise, nobody seems to criticize the empirical work of pre-modern scientists for their seeming lack of concern about the need to repeat their observations often enough to attain statistically meaningful sample sizes.

How then can we make sense of

what happens in courtrooms, or in astronomy or biology – or in case studies? One way to speak statistically about domains such as astronomy is to argue that they confront zero or near-zero sample variability – the members of the population are so similar on the dimensions of interest that the informational value of additional observations approaches zero. To the uninitiated, an *a priori* assumption of zero sample variability is no more or no less plausible than an assumption of some arbitrarily large sample variability. If observations are costly and sample variability is believed to be quite low, then the case for more observations is hardly self-evident. However, it is probably not wise to proceed very far in political science on the assumption that sample variability can be neglected.

James Fearon has argued all causal inferences are statistically based. Yet Fearon himself provides a riposte to this contention in his discussion of what he terms “counterfactual” explanations:

[S]upport for a causal hypothesis in the counterfactual strategy comes from arguments [emphasis in original] about what would have happened. These arguments are made credible (1) by invoking general principles, theories, laws, or regularities distinct from the hypotheses being tested; and (2) by drawing on knowledge of historical facts relevant to a counterfactual scenario.

What Fearon offers is a strategy for constructing a non-statistical basis for causal inferences. However, if one can support causal inferences by means of arguments of the sort that Fearon mentions, then there is no need for counterfactual speculation. One can just move directly from the arguments to the conclusions about causal processes operating in the case, without any need to construct counterfactuals. Fearon’s strategy is always available, whether one is interested in constructing counterfactuals or not. (However, case study researchers might have other rea-

sons to be interested in counterfactuals.)

As applied to a setting such as a trial or a case study, two types of arguments can be mustered in support of causal conclusions. The first are causal claims that are so uncontroversial that they operate essentially as primitive terms. If the jury views an undoctored videotape in which a suspect is seen pointing a gun at the victim and pulling the trigger, and the victim then is seen to collapse with a gaping hole in his forehead, it reaches conclusions about the cause of the victim’s death and the intent of the suspect to shoot the victim that are highly certain. Barring the sort of exotic circumstances that a philosopher or a mystery writer might invoke (e.g., the victim died of a brain aneurysm just before the bullet struck, or the gun was loaded with blank cartridges and the fatal shot was fired by someone else) the assessment of causation is unproblematic. Even if exotic circumstances are present, a sufficiently diligent search has a good chance to uncover them, as any reader of detective fiction knows.

A second type of causal claim is weaker: It is the “circumstantial evidence” so often used by writers of murder mysteries. An observation may be consistent with several different hypotheses about the identity of the killer, and rule out few suspects. No one observation establishes the identity of the killer, but the detective’s background knowledge, in conjunction with a series of observations, provides the basis for judgements that generate or eliminate suspects. As the story progresses, we are usually presented with several instances in which “leads” (i.e. hypotheses arising from data) turn out to be “dead ends” (i.e. are falsified by new observations). Sometimes an old lead is revived when still more new observations suggest that previous observations were interpreted incorrectly, measures or estimates were mistaken, or low probability events (coincidences) occurred. Typically, the detective constructs a chronology of the actions of the relevant actors in which the central concern is with the timing of events and the assessment of who pos-

sessed what information at what time. This tracing of the causal process leads to the archetypal final scene: All the characters and the evidence are brought together and the brilliant detective not only supplies the results of the final observation that eliminates all but one suspect, but then proceeds to explain how the observations fit together to produce a consistent and accurate causal explanation of events. Rival theories are assessed and disposed of, generally by showing that they are not successful in accounting for all the observations. The suspect may attempt to argue that it is all a coincidence, but the detective knows that someone has to be the killer, and that the evidence against the suspect is so much stronger than the evidence against anybody else that one can conclude beyond a reasonable doubt that the suspect should be arrested.

This is what Wesley Salmon terms an “ontic” explanation. Although it rests on a foundation of observed regularities, the regularities themselves are only the basis for an explanation. The explanation provides an answer to a “why” or “how” question by providing mechanisms of (probabilistic) cause and effect:

The aim of a scientific explanation, according to the ontic conception, is to fit the event-to-be-explained into a discernible pattern. This pattern is constituted by regularities in nature – regularities which we often regard as laws of nature. ...

It should be immediately evident, however, that mere subsumption under laws – mere fitting of events into regular patterns – has little, if any, explanatory force. ... I cited the pre-Newtonian knowledge of the relationship of the tides to the position and phase of the moon as a prime historical example of subsumption of natural phenomena under regularities that was totally lacking in explanatory value. It was only when the Newtonian explanation of that

regularity in terms of the law of gravitation became available that anyone could maintain plausibly that the tides had been explained. The obvious moral to be drawn from the example, and many others as well, is that some regularities have explanatory power, while others constitute precisely the kinds of natural phenomena that demand explanation.

... To provide an explanation of a particular event is to identify the cause and, in many cases at least, to exhibit the causal relation between this cause and the event-to-be-explained. ...

The ontic conception is a *more* demanding standard than the following common statistical strategy in political science: (1) Positing a series of bivariate functional relationships between a dependent variable and various independent variables, rooted perhaps in intuition or in expectations formed as a result of prior research; (2) demonstrating statistical regularities in a set of observations; (3) claiming to have a satisfactory explanation of variation in the dependent variable because there is an adequate *statistical* accounting of covariation. From the ontic perspective, we do not have an adequate explanation of the phenomenon under study until we can say why the model works.

Equipped with this understanding of explanation, we can now make sense of Rogowski’s point that one case sometimes has an impact on theorizing that is way out of proportion to its status as non-quantitative, low-*n* “observation.” It is indeed difficult to understand such situations from the standpoint of a statistically based view of cases.

One way to understand the importance of a single case is to note that when the existence of a phenomenon is in question, only one case is needed to establish it. Case studies sometimes do just that. However, if something occurs only once, is it important *statistically*? King, Keohane and Verba, for example, describe Lijphart’s case study on political

cleavages in the Netherlands as “the case that broke the camel’s back.” For that to be so, the statistical camel would already have to be under a great deal of strain due to the accumulation of previous anomalous findings. But no other anomalous findings are mentioned. KKV also note that there had been many previous studies of the relation between cleavages and democracy. If so, the mystery of why this one study should have such an impact only deepens. Unless one believes that this particular prediction failure is especially threatening to the previous pluralist theory, the presence of many previous studies that found the predicted association between cleavage structure and democracy would provide all the more reason to write off Lijphart’s case study as an outlier. No statistical model is rejected because it fails to predict only one case, and the influence of any one case on judgments or computations about the true underlying distribution is a decreasing function of sample size – so more previous case studies would imply that Lijphart’s study would matter *less*. Unless the sample is quite small, then adding just one “observation” (assuming for the moment that a case study is just *an* observation) is going to make very little difference. And, from a conventional statistical standpoint, small samples are simply unreliable bases for inferences – whether one adds one additional case or not.

If one accepts that the Lijphart study had a pronounced impact on theorizing in comparative politics, and if one views this impact as legitimate and proper, there is no way to rationalize this via statistical thinking. Rogowski’s original suggestion for how to understand this situation – as an example of a clear theory being confronted with a clear outlier – is a step in the right direction. But if that were all that were happening, one would simply be presented with an unusually strong anomalous finding, to which one could respond in a large variety of ways.

If a case study can succeed in explaining *why* a case is an outlier by identifying causal mechanisms that were hitherto overlooked, it will have a much

more pronounced impact. It is not the fact that the old theory is strongly disconfirmed that makes a single case study so important; rather, it is its provision of new causal mechanisms in empirical accounts that fit the data at least once.

If this provides an appropriate framework of assumptions to use in assessing case studies, then what are appropriate criteria for evaluating the quality of cases? It is high time for exponents of case study approaches to address this question without attempting to force case studies into the Procrustean bed of logical positivism and classical statistical theory.

References

- Fearon, James D. (1991) "Counterfactuals and hypothesis testing in political science." *World Politics* 43: 176.
- King, Gary, Robert O. Keohane and Sydney Verba (1994) *Designing Social Inquiry: Scientific Inference in Quantitative Research*. Princeton: Princeton University Press.
- King, Gary, Robert O. Keohane and Sydney Verba (1995) "The importance of research design in political science." *American Political Science Review* 89: 477.
- McKeown, Timothy J. (1997) "Case studies and the statistical worldview," ms. under review.
- Rogowski, Ronald. 1995. "The role of theory and anomaly in social-scientific inference." *American Political Science Review* 89: 468.
- Salmon, Wesley C. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press, 121-122.
- Toulmin, Stephen E. (1972) *Human Understanding*. Princeton: Princeton University Press.

Mechanisms and Cases in Comparative Studies

Roger Petersen
Washington University
rpeterse@wuecon.wustl.edu

John Bowen
Washington University
jbowen@artsci.wustl.edu

Over the past few years, the pages of the *Newsletter* of the APSA's Organized Section in Comparative Politics have often called for a diffusion of ideas across existing intellectual divisions. Kenneth Shepsle and Barry Weingast urged comparativists to learn from the field of American Politics (Winter 1994); David Laitin, among others, asked for balance and interaction between area studies and deductive theory (Summer 1993; Winter 1995); Robert Bates argued that new emerging forms of analysis in comparative politics will suffer if not supported by the *verstehen* of ethnographers and historians (Winter 1996).

At Washington University, under the auspices of the Committee on Social Thought and Analysis, we tried to meet the worthy goals of integration and diffusion by bringing together a set of political scientists and anthropologists to discuss a variety of approaches to comparative studies. This short note highlights some of the outcomes of these meetings and previews the forthcoming volume *Critical Comparisons in Politics and Culture* (Cambridge University Press) that resulted from them.

In design as well as in presentation, the volume is inductive, bottom-up, case-based, rather than deductive, prescriptive, law-giving. This inductive approach, however, does not prevent us from drawing some strong conclusions. Taken as a whole, the volume supports comparative work that 1) incorporates ethnography 2) specifies and isolates micro-level mechanisms and processes, and 3) employs small-*n* comparison in a variety of ways. All of the political scientists in the volume provide examples of small-*n* comparison. While their designs differ, they all incorporate detailed knowledge of

cases and design their research so as to isolate causal mechanisms.

Furthermore, the volume as a whole illustrates a rich diversity of comparative techniques. It offers the reader a set of examples to ponder, argue with, and perhaps draw from in planning comparative components for their own research. Such diversity suggests that there is not one common underlying logic of comparison, as some political scientists have argued, but rather a set of complex choices entailing trade-offs.

The Different Emphases of Anthropology and Political Science

Not surprisingly, a wide difference in emphasis on the importance of generality divides the fields. Choices regarding the type of information and the form of analysis follow from this fundamental difference. While the political scientists value the construction of deductive models that could explain more than one case, the anthropologists are primarily concerned with validity, defined here as the degree to which an account picks up processes, ideas, or relationships that are indeed present in the world. This dissimilarity was enhanced by the more specific methodologies employed by the participants: all of the political scientists (Barbara Geddes, Miriam Golden, David Laitin, Margaret Levi, and Roger Petersen) draw on rational choice theory to varying extents, while the anthropologists (Fredrik Barth, John Bowen, Allen Johnson, Greg Urban) are concerned with the culturally specific.

Despite this difference in emphasis, there emerged some common ground that suggests some areas for developing dialogue, where further diffusion of ideas might be most productive.

The Importance of Ethnography

First of all, the political scientists agree with the anthropologists on the importance and utility of ethnography or detailed historical analysis. Miriam Golden's and Margaret Levi's chapters vividly illustrate this point. Miriam Golden's chapter analyzes puzzling trade union disputes. After years of studying

trade unions in Italy, she wondered why unions appeared to call strikes that were virtually unwinnable, in that the stated goal of the strike, preventing job loss, clearly could not be reached. Why this apparently irrational behavior? She then compared decisions to call strikes in several industrialized nations, and found a second anomaly. Although union leaders say they strike to prevent downsizing, they do not seem to respond more forcefully when more jobs are at risk. She concluded that the real motivation behind strikes was protecting the union itself. The Italian strikes were called to pursue this basic goal; they failed because leaders overestimated employees' willingness to follow their strike call. Both issues involve knowledge of the preferences of key actors, and how these preferences drive actions during a strategic process.

Golden's chapter powerfully demonstrates how rational choice approaches can rely on fieldwork to establish preferences. Her argument is convincing precisely because she attends to the details of process in each of her cases: what leaders and followers knew, how they assessed their chances, what happened after the strike call. Her argument gains further plausibility through her use of comparison. Golden develops maximum diversity by studying labor action in multiple industries across four states (England, Italy, United States, and Japan). Developing the benefits of a "most different" design, Golden identifies similar micro-level interactions despite the differences in environment.

Margaret Levi's chapter specifies the mechanisms that lead individuals to enlist (or not enlist) in armies. Levi shows that these micro-level mechanisms are produced within macro-level pathways. More specifically, Levi's model connects mechanisms of trust, ethical reciprocity, and calculation of cost of compliance to the nature and development of state-level institutions. In turn, explaining how these pathways formed and how they constrain individual action requires historically detailed analytical narratives that provide an understanding of institutions and cul-

ture. Bates has previously summarized the nature and requirements of analytical narrative in the pages of this newsletter: "Cultures are distinguished by their distinctive institutions. One of the major innovations in our discipline has been the creation of the tools with which to analyze institutions . . . The use of such methods requires precisely the kinds of data gathered by ethnographers, historians, and students of culture. It requires knowledge of sequence, perceptions, beliefs, expectations, and understandings" (Winter 1996). Levi's chapter is an extended illustration of this statement and suggests ways in which the knowledge and experience of anthropologists might augment or extend the comparative methods used by political scientists.

Mechanisms and Processes

A second area of general agreement among the participants was on the importance of discovering mechanisms and focusing on processes. We use comparisons not for their own sake, but because they allow us to better understand processes and mechanisms, the how and why of social phenomena. Mechanisms are specific patterns of action which explain individual acts and events; when linked together they form a process. Within the volume, the types of mechanisms that are sought vary to significant degrees, as shown by the chapters by Barbara Geddes and David Laitin.

Geddes's chapter specifies the conditions under which individual politicians, all presumably seeking to maximize electoral support, choose to support a specific policy. In her case, she sought to explain why Latin American politicians choose to support civil service reform in some instances and not in others. From her knowledge of Brazilian politics and society, Geddes developed a model that not only applied to Brazil but also was capable of generating testable hypotheses about a wider set of cases. Here, the mechanism is a carefully specified empirical context that explains the action of a rational (optimizing) individual. One benefit of rational choice models is that they explicitly define the cases in

which the mechanism should be triggered. Geddes used comparison of several Latin American countries to test the model's ability to predict.

David Laitin's chapter also draws on the concept of mechanism, but in an expanded sense. In order to explain the use of violence in some cases of national revival but not others, Laitin builds an explanatory model whose critical mechanism is the tipping point at which sufficiently many people participate in the movement to make the costs of participation drop. The model, however, specifies several mechanisms (reasons for individual action) that are not rational, in addition to those that are. Laitin includes "the tyranny of sunk costs" and "the culture of violence" as important forces that drive individual action and that are relevant to explaining variation in outcomes. Laitin uses two sets of comparisons within his study: one is most similar design (Catalonia vs. Basque Country) used to isolate mechanisms; the second employs a most different design (Georgia vs. Ukraine) to test the model's generality.

Anthropologists also search for fine-grained reasons to explain individual action. Their search, though, is likely to be less restricted. "Rational" mechanisms are not privileged; the desire to link the mechanism to specific empirical contexts may not be as strong; the desire to build a formal model that predicts when a mechanism will be "triggered" is generally absent. Yet the fact that both anthropologists and political scientists are searching for fine-grained forces driving individual action provides common ground for discussion. Indeed, without this common element, there would have been little affinity between this particular grouping of anthropologists and political scientists.

In the Introduction of *Critical Comparisons*, we provide a broad definition of mechanism, one that we hope is capable of applying to a wide range of settings and able to cross disciplines. Following Elster (1997), mechanisms are intended to apply over a wide range of settings, and often include psychological

predispositions. For example, someone might continue to keep and repair an old automobile despite the likelihood of additional costly repairs because he or she figures a lot has already been invested in the car. This mechanism, the “tyranny of sunk costs,” may also keep spouses together who would otherwise separate because they cannot accept the fact that investments in the relationship have been in vain. This mechanism is both broadly applicable to a wide variety of cases (cars and spouses—and also is used in Laitin’s model of national violence), and specific in that it can be used to explain why a particular event occurs.

This type of explanation via a mechanism does not, however, seek a high degree of predictive power, nor does it aim at the creation of general laws. Sometimes spouses do break up, and other mechanisms (“the grass is greener”, for instance) may be at work. “If *p* then sometimes *q*” is the closest to a prediction that can be made within this explanatory framework. This methodological choice differs from seeking predictive power through the use of a variable approach. It also serves to bring the anthropologists in the collection closer to the mechanism-seeking political scientists. While the mechanisms of the political scientists in the volume may be more tightly connected to assumptions of rationality underlying rational choice theory, both groups, through the search for mechanisms, emphasize developing a fine-grained understanding of particular processes rather than generating simplified propositions about the general relationship among variables. The importance of specifying mechanism and process again points to the value of case studies. Without the knowledge of detail and context provided by case study methods, mechanisms and processes cannot be convincingly identified.

The possibilities for expanded dialogue across the two disciplines through this expanded sense of mechanism can be seen more specifically in the chapters by Fredrik Barth, David Laitin, and Greg Urban. Barth, like Laitin, attempts to pinpoint psycho-social mechanisms, in

his case, the mechanisms causing variation in New Guinea initiation rituals. Both Barth and Laitin combine comparative method with independent modeling of explanations. However, Barth’s emphasis is on developing models to account for fields of variation, while Laitin’s priority is to develop variation to help build and validate models. Despite this difference, the work of both scholars shows how deductive models develop insights into processes that could not be gained inductively through use of controlled comparison alone.

Greg Urban’s chapter argues that a focus on the processes of transmission of cultural objects might often be more fruitful than a cross-cultural comparison of the objects themselves. It is these pathways of transmission that connect the object to different cultures and set up possibilities for comparison. In effect, Urban identifies mechanisms of diffusion that should be studied in their own right. Such mechanisms have broad applicability for political scientists. The objects of study in political science, such as Golden’s strikes or the 1989 demonstrations in Eastern Europe, possess meaning that is definitely transmitted across state and cultural lines. As Urban cautions, the social scientist must be aware of how these objects and their meaning may change during the process of transmission from place to place.

Small-N Comparison

These two points, the importance of ethnography and of mechanism and process, lead to a third commonality: the use of small-*n* comparisons. Should small-*n* comparison be used as a “stage” in comparative politics? Political scientists generally advocate small-*n* comparisons in the absence of strong theory or in the presence of a small universe of cases when the theory might apply. In other instances, political scientists often prefer large-*n* statistical studies that are able to determine whether the observable implications of a theory are occurring in a non-random way. Going back to Lijphart’s seminal 1971 article, such a view posits a division of labor among

practitioners of qualitative and quantitative methods. Lijphart wrote (1971:685): “If at all possible one should use the statistical (or perhaps even the experimental) method instead of the weaker comparative method.” The strength of small-scale or “small-*n*” comparisons, Lijphart continued, lay in their ability to help create coherent hypotheses in a “first stage” of research. A statistical “second stage” would test these hypotheses “in as large a sample as possible”.

While the political scientists in our group probably feel that the number of cases, and, relatedly, whether quantitative methods are appropriate, depends upon the kind of question and evidence at hand, the ability of small-*n* comparisons to clarify mechanism and process and to combine the benefits of ethnography and strategic analysis mean that such comparisons need not be relegated to a “stage” in the research process (see Collier’s letter in this issue for an expanded discussion of this point). Above, we have mentioned how five political scientists have used small-*n* comparison for different reasons and at different stages of research and data collection. These scholars show that small-*n* comparisons can be used to challenge existing theories and test new ones, that they can be used when the universe of possible cases is small and when that universe is large.

While we would not claim that small-*n* comparisons are always more appropriate, we (the editors) do feel comfortable advocating small-*n* studies for a wide range of issues, and stress the benefits they produce in handling the complexity of social phenomena.

One Logic of Comparison?

Despite fragmentation in actual practice, there is a political science tradition of attempting to delineate one fundamental logic that underlies all comparative study, both quantitative and qualitative, and perhaps all of social science. In their influential 1970 work, Adam Przeworski and Henry Teune (1970:86) conclude: “Although the phenomena under consideration vary from discipline to

discipline, the logic of scientific inquiry is the same for all social sciences. As the theories explaining social events become general, the explanations of particular events will cut across presently accepted borders of particular disciplines.” In another influential book published nearly two and a half decades later, Gary King, Robert Keohane, and Sidney Verba (1994:4) write: “A major purpose of this book is to show that the differences between the quantitative and qualitative traditions are only stylistic and are methodologically and substantively unimportant. All good research can be understood—indeed, is best understood—to derive from the same underlying logic of inference.”

Our approach differs. While we applaud the search for common ground, we believe the differences among the disciplines are more than a matter of style. Certainly, the prevailing goals vary among fields, if not the respective logics. Rather than trying to convince social science practitioners that there is one underlying logic, or developing a new synthesis, we believe that interdisciplinary progress might best be made by presenting choices and trade-offs made in the course of quite distinct research projects. We believe that knowing a wider range of possible ways of comparing will both help individual researchers in their own work and help build bridges across disciplines.

References

Elster, Jon (1998), *Alchemies of the Mind: Studies in Rationality and the Emotions*, forthcoming Cambridge University Press.

Lijphart, Arend, “Comparative Politics and the Comparative Method,” *American Political Science Review* 65 (Sept. 1971) pp. 682-693.

King, Gary and Robert O. Keohane and Sidney Verba, (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton: Princeton University Press.

Przeworski, Adam and Teune, Henry, (1970), *Logic of Comparative So-*

cial Inquiry, New York: Wiley-Interscience.

Comparative Methodology, Fuzzy Sets, and the Study of Sufficient Causes

Charles C. Ragin
Northwestern University
University of Oslo, Norway
cragin@casbah.acns.nwu.edu

Overview

This essay previews some of the arguments I present in my forthcoming book, *Fuzzy Social Science*. The term “fuzzy” is used not as a synonym for “loose” or “haphazard,” the everyday usage, but in reference to fuzzy set theory (Zadeh 1965), a framework that permits precise operationalization of theoretical categories. I begin by evaluating a common form small-*n* inquiry, the study of the causal conditions shared by multiple instances of the same outcome, and briefly review some of the abuse this design has recently received from quantitatively oriented social scientists. I then argue that the primary weakness of this design is *not* its failure to live up to the standards of conventional quantitative social science, but its limited approach to causation. Specifically, this design is capable only of identifying necessary-but-not-sufficient causes and assumes implicitly that such causes exist. I sketch a more fruitful approach to causal complexity, the study of the *sufficiency* of causal combinations.

One obstacle to implementing this (and most other) case-oriented designs is the difficulty of making “yes/no” assignments of cases. Dichotomizing always leaves researchers open to the charge that they have selected cut-off values that favor one argument over another. Conventionally, the problem of dichotomizing is solved by applying correlational methods to interval-scale variables. I argue (1) that in most research situations fuzzy sets are more useful than interval-scale variables because fuzzy sets offer more faithful representations of theoretical concepts, and (2) that the

use of correlational techniques is directly at odds with the assessment of causal sufficiency. Correlation counts as “error” many cases that are perfectly compatible with a causal argument emphasizing conditions that are sufficient but not necessary for an outcome.

The Analytic Value of Case-Oriented Research

One common approach to small-*n* design is to examine three or more cases in which a given outcome occurs. Usually, this design involves studying commonalities across a relatively small number of countries (or other macro-level units) displaying roughly similar outcomes. For example, a researcher might study several “anti-neocolonial revolutions.” Research of this type usually has a dual focus. The first and most basic task is to clarify the outcome. For example, an investigator might seek to answer the question: What, exactly, is an anti-neocolonial revolution? The researcher pinpoints similarities across instances of the outcome and contrasts these commonalities with what is known about different-but-related outcomes—other types of revolutions, for example. Assuming this first task—defining the outcome with adequate precision—can be accomplished, the researcher then proceeds to the second, more decisive task—examining relevant causal conditions. Here, the researcher uses existing theory and substantive knowledge to pinpoint the causes of the outcome in question. Ideally, instances of the outcome will agree on a number of causal conditions that make theoretical and substantive sense as jointly necessary conditions. Sometimes fancy theoretical footwork is required to establish as commensurate causally equivalent but empirically different causes, and often researchers must conclude that there are several different sets of causal conditions capable of producing roughly the same outcome. Still, the search for shared causal conditions is the usual starting point.

From the perspective of conventional quantitative social science, the research design just sketched is ludicrous. First

and most obvious is the simple fact that the number of cases is usually too small to permit any sophisticated form of data analysis. Second and much more damning, at first glance, is the fact that the design proceeds from selecting on the dependent variable (studying multiple instances of “the same thing”) to searching for causal conditions that are invariant across cases (causal “variables” that do not vary). Thanks to the diligent efforts of King et al. (1994) and other scholars, many students of political science are now convinced that this research design forces researchers to commit unforgivable sins. An example of this conviction is the growing tendency, at least in the dissertation proposals I have read over the last several years, for beginning case-oriented researchers to select cases that display “high, medium, and low” values on an outcome, conceived as representing the range of variation of what is presumed to be an underlying interval-level variable. Unfortunately, the usual pattern is for this design to culminate in the production of three or more vaguely connected case studies.

In fact, however, the design just sketched—studying causal conditions shared by instances of “the same thing”—is a perfectly reasonable way to conduct social science. In a recent article (Ragin 1997) I address common criticisms of this and related forms of case-oriented comparative research and then proceed to “turn the tables” and show how the *practical* concerns of case-oriented researchers cannot be addressed with conventional quantitative methods. Indeed, the quantitative critiques of case-oriented designs argue that at the *outset* of an investigation the case-oriented researcher should have (1) a well-articulated and testable theory, (2) a well-defined and delimited population of relevant observations, and (3) a clear specification of the key features of the outcome under investigation. These common preconditions for conventional forms of quantitative analysis are rarely met at the outset of most case-oriented research.

While routinely scorned by quanti-

tative researchers, the study of the causal conditions shared by multiple instances of the same outcome is especially useful for evaluating *necessary* conditions. A necessary condition must be present for the outcome in question to occur. To assess necessity, the researcher works backwards from instances of the outcome to the identification of relevant causal conditions shared by these instances. All, or virtually all, instances of the outcome should be preceded by the same cause or set of causes. If the researcher successfully identifies relevant shared causal conditions, then they can be portrayed as jointly necessary conditions, when viewed through the lens of the researcher’s theoretical framework. Alternatively, the researcher might wish to use a failed search for common causal conditions to reject hypothesized necessary conditions and thus challenge existing theory.

Necessity Versus Sufficiency

Of course, necessary conditions are only rarely *sufficient*. Even if it can be shown, for example, that all known anti-neocolonial revolutions were preceded by a given set of causal conditions, it would be erroneous to conclude that whenever these causal conditions converge, an anti-neocolonial revolution results. To assess the sufficiency of a cause or causal combination, the researcher must determine whether or not the cause or combination of causes in question always, or virtually always, produces the outcome in question. Evidence that there are instances of the cause or causal combination *not* followed by the outcome challenges the researcher’s claim that the cause or causal combina-

tion is sufficient. The assessment of sufficiency, therefore, involves searching for cases that are similar with respect to the cause or causal combination in question and then assessing whether or not they agree in the outcome. For illustration, consider Table 1, which shows a crosstabulation of the presence/absence of a cause against the presence/absence of an effect. The test of *necessity* involves the first row (cells 1 and 2). To pass the test of necessity, cell 1 should be empty, or virtually empty. The assessment of *sufficiency* involves the second column. To pass the test of sufficiency, cell 4 should be empty, or virtually empty. Cell 3 is not directly relevant to either assessment.

In many fields of social science, the assessment of necessity is very important. However, it is essential to recognize that designs that work backwards from multiple instances of an outcome to shared preconditions (1) cannot address sufficiency and thus often will fail to identify *decisive* causes, and (2) makes restrictive assumptions about causation, namely, that theoretically relevant necessary conditions *exist*. I contend that the search for necessary causes, while often useful, constricts case-oriented research in unproductive ways. As I argue in *The Comparative Method* (1987) and elsewhere, in case-oriented research it is much more productive to allow for the possibility that a given outcome results from a variety of different combinations of theoretically relevant causal conditions. The investigator may find that these different combinations share one or more causal conditions. If so, then he or she may argue that the shared conditions are necessary conditions (see Ragin

	Cause absent	Cause present
Outcome present	1. key cell for assessing necessity; cell should be empty	2. relevant to both assessments
Outcome absent	3. not directly relevant to either assessment	4. key cell for assessing sufficiency; cell should be empty

1987:99-101). But this process of examining multiple causal conjunctures is very different from making the assumption at the outset of an investigation that necessary causes exist.

As an alternative to searching for necessary-but-not-sufficient causal conditions, researchers should work forward from causal conditions, especially combinations of conditions, and assess their sufficiency. Basically, this alternate design involves “selecting on the *independent* variables” (i.e., looking at cases with a specific combination of values) and assessing whether or not these cases display the same outcome. For illustration, suppose there are several combinations of conditions that produce anti-neocolonial revolutions. Because there are several sufficient combinations, the analysis of necessity would show that no combination is necessary. However, the analysis of the sufficiency of each combination would show that each is sufficient. The general pattern is illustrated in Table 2. In the analysis of any of the sufficient causal combinations, there would be cases in cell 1 because there is plural causation. However, every relevant combination would pass the test of sufficiency. In each test, cell 4 would be void or virtually void of cases, while cell 2 would contain instances of anti-neocolonial revolution explained by the combination in question.

Table 2 demonstrates that it is possible to assess the sufficiency of causal combinations one at a time, in isolation from one another. This conclusion is important because of its implications for the study of causal complexity. If, as I have argued, we live in a world of great causal complexity, then a common pattern will be for outcomes to result from different combinations of causal conditions. While it might seem that causation this complex would befuddle analytic social science, it is clear from the example just presented that the analysis of the sufficiency of causal combinations can proceed in a straightforward manner.

The Problem of Fit

Unfortunately, it is often difficult to

Table 2: Assessing Multiple Causal Combinations		
	First causal combination absent	First causal combination present
Outcome present	1. cases explained by the other causal combinations	2. cases explained by first causal combination
Outcome absent	3. not directly relevant to the assessment of sufficiency	4. no cases (or virtually no cases) here

conduct analyses like the one just described. While there are several obstacles worthy of discussion, I focus on only one here—the simple fact that social phenomena are not as “crisp” as we would like. Social phenomena lack crispness when their membership in the sets that social scientists use is partial or incomplete. For example, political scientists like to make statements about “democracies.” However, countries differ in the degree to which they are members of this set. The issue is not simply one of definition, but the fact that even when armed with a very precise definition, a researcher will find it difficult to make crisp “yes/no” assignments. Is Mexico a democracy? Is Russia? Empirical cases vary greatly in the degree to which they “fit” the categories that derive from the concepts that social scientists use.

The issue of crispness (and its opposite, fuzziness) is crucially important to the assessment of the sufficiency of causes, as just sketched. To assess the sufficiency of a causal combination, the researcher must first select cases displaying that combination. But what if cases vary in the degree to which they display a particular causal combination? Which ones should he or she select? Likewise, after selecting relevant cases, the researcher must assess whether or not these cases display the outcome in question. But what if cases vary in the degree to which they express the outcome? At first glance, this task appears to be a job for our dear old friend, correlational analysis of interval-scale variables. However, as I show subsequently, correlational analysis is *not* appropriate

for the examination of the *sufficiency* of causal conditions. Instead, researchers should represent their theoretical concepts as fuzzy sets and examine the set-theoretic relationship between causal conditions and outcomes. Before providing this demonstration, I first offer a very brief overview of the nature of fuzzy sets.

Fuzzy sets extend conventional logic by permitting membership scores in sets to take values in the interval between 0 and 1. Conventional logic, by contrast, permits only the scores of 0 (nonmembership) and 1 (membership). For example, with fuzzy sets a person might receive a membership score of .75 in the set of European Americans and a score of .9 in the set of heterosexual males. The basic idea is to permit the *scaling* of membership scores and thus allow partial or “fuzzy” membership in sets. A membership score of 1 indicates full membership in a set; scores close to 1 indicate strong but partial membership in a set; scores less than .5 but greater than 0 indicate that objects are more “out” than “in” a set, but still weak members of the set; a score of 0 indicates full nonmembership in the set. Thus, fuzzy sets combine qualitative and quantitative assessment: 1 and 0 are qualitative assignments (“fully in” and “fully out,” respectively); values between 0 and 1 indicate degrees of membership.

For illustration, consider fuzzy membership in the set of “rich countries.” A conventional variable like GNP per capita offers a good starting point for assessing membership in this set, but the translation of this variable to fuzzy membership scores is neither automatic nor

mechanical. It would be a serious mistake, for instance, to score the poorest country 0, the richest country 1, and then to array all the other countries between 0 and 1. Instead, the first task in this translation would be to specify three important qualitative anchors: the point at which full membership is reached (i.e., definitely a rich country, membership score = 1), the point at which full non-membership is reached (i.e., definitely not a rich country, membership score = 0), and the point of maximum ambiguity in whether a country is “more in” or “more out” of the set of rich countries (a membership score of .5, the cross-over point). When specifying these qualitative anchors, the investigator must present an explicit rationale for each breakpoint.

One possible translation of GNP per capita values to fuzzy membership scores for the set of rich countries would use the following qualitative breakpoints: Countries with GNP per capita values of \$1,000 or less are definitely out of the set of rich countries; countries with values between \$1,000 and \$5,000 are more out than in, but not fully out; \$5,000 is the cross-over point, where there is maximum ambiguity in whether a country is more in or more out; countries with values between \$5,000 and \$10,000 are more in than out of the set of rich countries, but not fully in; and countries with values of \$10,000 or more are fully in the set of rich countries. GNP per capita values below \$1,000 are compacted into the fuzzy score of 0 because these cases are all equally “out” of the set of rich countries. Likewise, GNP per capita values greater than \$10,000 are compacted into the fuzzy score of 1 because these cases are all equally “in” the set of rich countries.

The use of qualitative anchors to identify key breakpoints on continua contrasts sharply with conventional social science, where the usual concern is to maximize the variation of all variables. Thus, the use of fuzzy sets challenges the implicit notion of much conventional work that all variation is meaningful. From the perspective of fuzzy logic, theoretic

cal concepts (e.g., “rich countries”) are paramount, and the central problem is to assess membership in such sets. Some regions of the range of a conventional variable used as an indicator of a concept may be irrelevant to the theoretical understanding of the phenomenon under investigation. From the perspective of fuzzy social science, to go from a theoretical formulation involving sets with fuzzy membership to correlations between variables containing substantial amounts of irrelevant variation simply adds error to ambiguity.

It is not possible in this short essay to offer a more detailed presentation of the many ways that fuzzy sets bring clarity to social scientific work. I return to my central focus, the examination of the sufficiency of causal conditions.

Fuzzy Sets and the Assessment of Sufficiency

Recall that a central obstacle to the assessment of causal sufficiency is the fact that it is difficult to make crisp “yes/no” decisions regarding whether or not cases display a causal combination and whether or not they display a particular outcome. As an alternative, researchers can assess cases’ fuzzy membership in the sets defined by the causal combination and the outcome. For example, suppose the investigator has reason to believe that anti-neocolonial revolutions occur when four conditions converge (e.g., foreign capital domination combined with multiple sovereignty combined with . . . etc.), and that this combination is only one of several combinations of conditions that spawn anti-neocolonial revolutions. The investigator can assess the fuzzy membership of relevant cases in these two sets (the set defined by the outcome and the set defined by the causal combination) and then examine the set-theoretic relationship between the two sets of scores.

It is very important to emphasize that this assessment should *not* be correlational. If the researcher is interested in assessing the sufficiency of one of many sufficient causal combinations, as in the present example, then there will be a

How to Subscribe

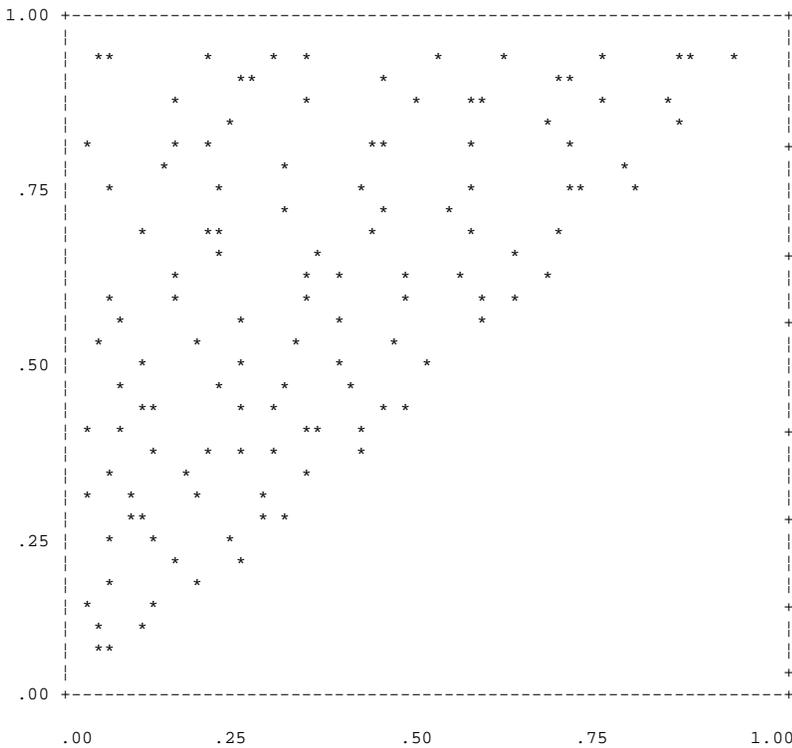
Subscriptions to the *APSA-CP Newsletter* are a benefit to members of the Organized Section in Comparative Politics of the American Political Science Association. To join the Section, check the appropriate box when joining the APSA or renewing your Association membership. Sections dues currently run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

*American Political
Science Association*
1527 New Hampshire Ave., NW
Washington, DC 20036
Telephone: (202) 483-2512
Facsimile: (202) 483-2657
Email: membership@apsa.com

substantial number of cases that have *strong* membership in the set of cases displaying the outcome but *weak* in the set of cases conforming to the combination of conditions in question. This pattern would obtain for the simple reason that there is more than one way to spawn an anti-neocolonial revolution. Indeed, if this researcher were to construct a scatter plot of “membership in the set of cases with anti-neocolonial revolutions” against “membership in the set of cases displaying one of the several combinations of causal conditions that spawn anti-neocolonial revolutions,” he or she would *expect* to find a roughly triangular pattern, as depicted in Figure 1. This figure shows, in effect, that the fuzzy set of cases displaying the causal combination is a subset of the fuzzy set displaying the outcome. Cases in the upper left-hand corner of this plot do *not* violate the argument that the causal combination is sufficient; only cases that fall well below the main diagonal contradict this argument. The lower right corner of this plot corresponds directly to cell 4 of Tables 1 and 2, the “forbidden” cell in the assessment of causal sufficiency.

Correlational analysis, by contrast, would treat many of these cases, espe-

Figure 1: Plot of Membership in the Set of Anti-Neocolonial Revolutions (Y Axis) Against Membership in the Set Defined by One of the Causal Combinations Sufficient for Anti-NeoColonial Revolution (X Axis)



cially those in the upper left corner, as errors, even though no case above the diagonal contradicts the argument that the causal combination in question is sufficient for the outcome. Thus, the rush to correlational analysis that is so common among social scientists could easily lead to the rejection of the clear demonstration of causal sufficiency depicted in Figure 1.

Conclusion

While there are many lessons to be learned from the examination of necessary causes, the advance of social scientific knowledge is best served when scholars make as few assumptions about causation as possible, especially at the outset of an investigation. When scholars assume maximum causal complexity—that different combinations of causes may produce the same outcome, they assume that no single cause is either necessary or sufficient. As I have shown, analytic social science is possible even when causal complexity is great. The analysis of causal complexity, in turn, is greatly facilitated by the use of fuzzy sets. This approach offers a pow-

erful way to assess the sufficiency of causal conditions, a task that is outside the domain of conventional correlational analysis.

References

- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles C. 1997. "Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research." *Comparative Social Research* 16:27-42.
- Ragin, Charles C. forthcoming. *Fuzzy Social Science*. Chicago: University of Chicago Press.
- Zadeh, Lofti A. 1965. "Fuzzy Sets." *Information Control* 8:338-53.

Historical Analysis and Causal Assessment in Comparative Research¹

John D. Stephens
University of North Carolina
jdsteph@unc.edu

As a scholar who was trained as a sociologist but then migrated more than ten years ago into political science, I certainly follow developments in our sister discipline more closely than most political scientists. I would like to take the opportunity provided by the editor of this newsletter to share some of my thoughts about a recent methodological debate in comparative sociology and its relationship to the exchange on qualitative methodology in political science generated by King, Keohane, and Verba's (KKV) *Designing Social Inquiry* and carried out at the 1994 APSA meetings and in the *APSR* in June 1995. In a parallel debate in the *British Journal of Sociology* in 1991, John Goldthorpe issued a major challenge to the practice of "grand historical sociology" exemplified by the works of Barrington Moore, Theda Skocpol, and Michael Mann, which subsequently led to an exchange in the pages of that journal.² Goldthorpe continued his criticism of comparative historical sociology in the most recent issue of *Comparative Social Research*, this time focusing criticism on Charles Ragin's *The Comparative Method* and my book with Dietrich Rueschemeyer and Evelyne Huber, *Capitalist Development and Democracy* (CDD). Here I would like to take up one issue that Rueschemeyer and I addressed in our response to Goldthorpe, which appeared in the same issue of the annual, as it is one which also points to a major flaw, in my view, in KKV's discussion of qualitative methodology, that is, their treatment of "causal mechanisms," "process tracing", and "historical analysis," all alternative labels for essentially the same research procedure. Of the four questions raised by Goldthorpe, three, the small-*n* problem, the Galton (or diffusion) problem, and the "black box" problem (the term we use

in the book for the problem of moving from correlation to causality) relate to this issue.³ Goldthorpe criticizes Ragin and CDD because he contends that we explicitly or implicitly argue that comparative historical research is methodologically superior to crossnational quantitative research, while he claims that with regard to three central problems it does not live up to the claim. In fact, we do not make blanket claims of superiority in the book, but, ironically, with regard to these three specific problems, we would make such a claim. Our response to all of them is a common one: Comparative historical analysis is actually superior to crossnational quantitative analysis, its only competitor in the study of these macro social developments, because it allows one to uncover the causal processes and thus to eliminate rival explanations for the phenomenon under study. In the short space I have here, I will develop our argument with regard to the small-*n* problem and then move on to consider KKV's arguments on this question. I will also briefly address the Galton and black box problems in passing.

The Small-N Problem

Goldthorpe is correct in arguing that both in quantitative variable-oriented analysis and in case-based comparative research the number of variables often exceeds the number of cases, making the testing of competing theories impossible. In that situation, one might find a number of different explanations supported equally well by the data with no way to distinguish among them, as Huber, Ragin, and I (1991) have empirically shown in the case of cross-national statistical research on the welfare state. Another example illustrates the problem in case-based comparative research. A recent collaborative research project on the breakdown of democracy in interwar Europe included more than 20 countries, all of the countries in Eastern and Western Europe, but nonetheless faced the same problem. Two of the collaborators, in undertaking an analysis with Ragin's Qualitative Comparative Analysis (QCA), identified over sixty characteristics which various theories had hy-

pothesized to be related to democratic collapse (Berg-Schlosser and De Meur 1994). It is not surprising that the authors could produce a number of different, and in some cases theoretically contradictory, solutions. QCA, as Ragin has developed it, is essentially a formalization of Mill's indirect method of difference or what Skocpol calls the "macro analytic" approach in comparative historical sociology. Thus, both comparative methods are apparently incapable of resolving the problem.

In some cases, both the comparative method and quantitative analysis provide a criterion, essentially the same criterion, for moving beyond this point.⁴ Unfortunately, this one criterion can lead to fallacious conclusions. In statistical analysis, when choosing between two or more regressions (or any other technique) of statistically equal explanatory power, the regression with the fewest variables is favored. To restate the same principle in slightly different terms, a single variable is favored over two competing variables with equal explanatory power. A similar assumption is often made in comparative analysis, an assumption which can be most clearly seen in QCA. Applying Occam's razor, QCA assumes that the solution with the fewest explanatory characteristics is the best. However, this may not identify the true causal variables. To take a hypothetical example, assume that in an array of cases a characteristic Y is the dependent variable of interest and there are two different paths to this outcome, A and B. Yet if all cases having Y also have characteristic C (because A and B cause C, or Y causes C, or by pure chance), then C rather than A and B will be identified as the explanation, both in statistical analysis and in QCA.

Other than this potentially misleading criterion, there are no criteria provided for by the logic of the comparative method or by quantitative methodology for choosing between solutions of equivalent statistical or logical power or for distinguishing spurious correlations from causal factors. However, it is misleading to reduce the comparative historical method to Mill's method of dif-

ference or Ragin's QCA. Even when it is applied to contemporary societies, the comparative historical method is also historical in that it involves tracing the historical process. By uncovering agency and historical sequence, one can eliminate some potential causal variables and strengthen the case for others.

Let me take an example from CDD to illustrate this, an example which at the same time shows how comparative historical analysis can address the Galton problem. We note that the correlation between democracy and British colonialism is robust. This statistical association has been given a diffusionist interpretation: British colonialism made a positive contribution to democratization in its colonies through the transfer of British governmental and representative institutions and the tutoring of the colonial people in the ways of British government. We did find evidence of this diffusion effect in the British settler colonies of North America and the Antipodes (p. 280); but in the West Indies, the historical record points to a different connection between British rule and democracy (Chapter VI, also see pp. 280-81). There the British colonial administration opposed suffrage extension, and only the white elites were "tutored" in the representative institutions. But, critically, we argued on the basis of the contrast with Central America, British colonialism did prevent the plantation elites from controlling the local state and responding to the labor rebellion of the 1930s with massive repression. Against the adamant opposition of that elite, the British colonial rulers responded with concessions which allowed for the growth of the party-union complexes rooted in the black middle and working classes, which formed the backbone of the later movement for democracy and independence. Thus, the narrative histories of these cases indicate that the robust statistical relation between British colonialism and democracy is produced only in part by diffusion. The interaction of class forces, state power, and colonial policy must be brought in to fully account for the statistical result.

King, Keohane, and Verba on "Causal Mechanisms" and "Process Tracing"

The claim that one can establish cause or at least strengthen the case for a causal factor or set of causal factors over another set by carrying out historical analysis has been contested by KKV in their recent book. They contend that historical analysis or process tracing, to use George and McKeown's terms for a similar analytic procedure, cannot establish cause. Conceptually, they contend that the sole way to establish cause is experimental control. In most of the problems we face in political science, hypothetically, the only way to do this would be to rerun history with everything but the experimental variable held constant. So to take an example from my current research, to establish that the Labor Party was responsible for the establishment of the national health service in New Zealand, KKV contend that one would have to rerun New Zealand history without Labor in power in the 1940s but with everything else about New Zealand history, and world history in so far as it had an effect on New Zealand, the same as it actually occurred. I think KKV are in principle correct on this account; this is the only way to definitively demonstrate cause. Thus, even if we can show that the Labor Party in New Zealand ran on a platform calling for a national health service; that their opponents, the Nationals, opposed it; that Labor won and that Labor implemented the reform; we still do not know for sure that the election of Labor was the causal factor. It could have been something else and rerunning history would reveal this.

Moreover, I think KKV are correct in asserting (p. 85 ff.) that uncovering a "causal mechanism" is not only not a substitute for the experimental method, it also is not necessary to establish cause. They address the contention that to establish causality one must "identify a list of causal links between the two variables," objecting that "there always exists in the social sciences an infinity of causal steps between any two links in the chain of causal mechanisms" and that such an "approach quickly leads to infinite regress and at no time does it

alone give a precise definition of causality for any one cause and one effect" (p. 86). This seems to me to be true: If we could rerun New Zealand history without Labor in power in the 1940s, we could dispense with tracing the historical process to help determine whether Labor's electoral victory was decisive. But we can't rerun history and given this, these observations about historical sequence and agency are highly relevant if not in establishing cause, then getting closer to it and in sifting out spurious factors. Suppose we had found that the Nationals did not oppose the national health service, or that it wasn't in Labor's platform, that a bureaucrat had designed the reform? Wouldn't this be relevant information for narrowing the range of potential causes?

Of course, KKV recognize that "re-running history" is not possible and thus that in political research cause cannot be definitively established. Since possible alternative causes cannot be eliminated by random assignment as they would be in an experiment (or KKV's counterfactuals), statistical control is the best practical alternative. In comparative research, this leads us back to the small-*n* problem discussed previously. KKV argue, and I would concur, as would Goldthorpe, that a research design in which one has more variables than observations is indeterminate (118 ff.). KKV's solution is to increase the number of observations (Chapter 6) and they approvingly cite "process tracing" as one way of doing this (p. 226). However, as Sidney Tarrow has pointed out in his contribution to the *APSR* forum, by assimilating process tracing "to their favorite goal of increasing the number of theoretically relevant observations" (p. 472), KKV depart fundamentally from George and McKeown's conception of the method. In the case of historical analysis, their conception would be the equivalent of time series analysis, that is, adding observations with measurements of the theoretically relevant dependent and independent variables at different points within the case (country) in question. This is very different from the procedure followed by a good

historian who is attempting to identify the causes of phenomena of interest in the case he or she is researching. In his contribution to the exchange with Goldthorpe, Jack Goldstone characterizes this procedure as "unraveling historical narrative" and draws an apt analogy between the research process of the historian attempt to establish the causes of an event and the activity of a police detective attempting to establish the cause of traffic accident.

Tarrow makes another point bearing on the role of historical analysis in establishing causality which dovetails nicely with arguments Rueschemeyer and I advance in our exchange with Goldthorpe. In discussing Putnam's *Making Democracy Work*, Tarrow points out that after Putnam and his collaborators had done countless elite and mass surveys establishing the vast differences between Italy's north and south, Putnam was then forced to turn to history in an attempt to impose a causal structure on the crossregional correlations he had found. Thus Putnam faced the same "black box" problem that we found in our study of democracy when surveying the results of crossnational quantitative studies: Entirely different theoretical accounts were consistent with the quantitative data and there was no way to uncover the causal processes without turning to comparative history. True, when one has strong theoretical reasons for assuming that one of two strongly correlated variables is the explanatory factor and the other the dependent variable,⁵ one may be willing to make the leap to positing cause. But note that if Putnam is correct, extant theory would have been wrong in this case as it suggests that the level of development determines the strength of the civic community and not the other way around.

This problem of causal inference is not an idiosyncrasy of social science, as one could make similar observations about natural science. This is most obvious in cases in which scientists are forced to rely on data collected on natural populations and thus the researcher has no control over which subjects are assigned to the experimental and con-

trol groups, as in most epidemiological studies. But even in situations in which true experiments are possible, one may be more confident in one's results if one attempts to fill in the causal mechanism, taking at least one step down that infinite regress which KKV warn us against. As Lakatos (1970) has pointed out, no perfect experiment is possible which subjects scientific theories to definitive refutation (or "verification"). Suppose, for example, that we are studying the effect of a toxic substance on cancer. We administer the agent on our randomly selected experimental group of rats and a placebo on a control group. We find, as expected, that our experimental group did have much higher rates of cancer. This could, of course, have been a chance happening, an event that would have happened only once in 5000 times. KKV, no doubt, would tell us that this is why we need to replicate the experiment. But the result might have been due to something systematic, correlated to how we conducted the experiment but totally unknown given the state of knowledge in the field, which is exactly Lakatos's point about the impossibility of designing the definitive experiment. But suppose we then supplemented our experiment with micro biological research dissecting the rats and examining the cells under an electron microscope and we found the toxic substance caused gene breakage and the mutated cells became cancerous. As a result, we would be much more confident that we had the correct causal factor.⁶

To summarize my argument, the tracing of the historical sequences is an important analytical tool in social science in establishing cause, or better said, in narrowing the range of possible causes in a social phenomenon of interest, whether it be one occurring in the distant past or in the contemporary world. Moreover, when the social scientist is faced with a small number of cases or the possibility of spurious correlation regardless of the number of cases, it is an essential tool. It is a fundamentally different analytical procedure from increasing the number of observations. Rather, it involves linking these obser-

vations together in an historical narrative and examining human agency to strengthen the case for a given explanation of the historical outcomes over rival explanations.

¹ I would like to thank Evelyne Huber, Dietrich Rueschemeyer, and Sidney Tarrow for comments on an earlier draft of this essay.

² In my view this exchange generated a lot of heat and not much light. A more measured and insightful response to one issue raised by Goldthorpe, how one chooses among conflicting historical accounts of the same event (e.g. Moore's selection of a class analytic account of the English civil war over others as the "true" account), by Ian Lustick appeared in the *APSR* last year.

³ The fourth question concerns the problem of developing and testing theories inductively. Again on this issue, Goldthorpe and KKV argue along similar lines. This is unquestionably an important issue but beyond the reach of this short article. Rueschemeyer and I do discuss this question in our response to Goldthorpe.

⁴ This is not to imply that the QCA and quantitative solutions are the same. The primary difference is that, rather than establishing associations between variables, QCA establishes associations between characteristics and does so in a way that leaves the links of a particular set of characteristics with a case transparent.

⁵ Or which of two or three independent variables is the causal factor or what combination of them, etc.

⁶ This is not, it seems to me, a strict parallel to historical analysis or process tracing. A strict parallel would be to observe (or have an historical record which records) the actual process of the toxic substance coming into contact with the gene and so on.

Ecological Inference and the Comparative Method¹

D. Stephen Voss
Harvard University
dsvoss@wjh.harvard.edu

David Lublin

University of South Carolina
lublin@garnet.cla.sc.edu

If any obstacle dominates the study of comparative political behavior, it is the difficulty of obtaining high-quality data. Stable, reliable survey organizations are rare outside of the industrialized world. Even within the more developed countries, few polls match the standards and diversity of questions demanded for academic work. Few surveys are implemented in a consistent manner across national boundaries. And commercial polls seldom produce meaningful geographical detail, even within the survey-rich United States.²

Of course, not all comparative questions lend themselves to mass surveys. Field work is usually the best way to explore behavioral hypotheses, because the researcher gains insights that might escape more remote scholars. Regardless of whether field work concentrates on masses or (as is more common) elites, though, it consumes precious funds. The costliness places a premium on efficiency and often forces regrettably tough choices; few are so flush that they can pursue every intellectual lead. Therefore, any source of cheap data on mass behavior would enrich the profession by allowing comparativists to supplement their field work.

The one data source already present in most countries, yet consistently neglected in the study of political behavior, is the state. Most national governments collect reasonably reliable statistical information on the populations they rule. Further, democratic governments often retain voting returns for years after elections take place – and even if a government does not bother, political activists or firms contracted to manage electoral technology may have done so. Researchers sometimes can collect such data from home, using the Internet or placing a telephone request (although we have heard of cases where travel and bribes were necessary to collect government statistics). Even leaving aside the cases when surveys are impossible, such as when the researcher has historical interests or thin

financial resources, piggybacking on government efforts may be better than performing all the legwork alone.

What explains the gap between this wealth of raw behavioral data and the sparse use to which it is put? The main obstacle is one of technique: both electoral and census data are grouped over geographical units, and the profession has lacked a reliable method for aggregate-data analysis. For this reason, the recent solution to the “ecological inference problem” posed by Harvard University methodologist Gary King should be an invaluable tool for expanding the comparative method into questions and data sources that until now have proved intractable.³

The Ecological Inference Problem

Census data usually are reported in tables, with the figures aggregated over geographical units such as provinces or electoral enumeration districts. Election returns also are collected and reported for discrete areal units. This convention is in some ways a valuable one; it retains meaningful geographic variation (usually lost from surveys), with enormous sample sizes for each region. Yet it also sacrifices the individual-level detail that is often the object of comparative inquiry.

Table 1 presents a typical situation. We wish to know the extent of francophone support for sovereignty in Quebec’s 1995 referendum, yet election returns are not reported by linguistic group and we do not have the individual-level data to compute such a breakdown ourselves. The empty cells in this table therefore reflect the missing information. The marginal totals, on the other hand, represent data typically available to a comparative researcher: overall voter preferences, and the number of potential voters from each grouping of interest.

If we only possessed such figures for Quebec as a whole, the internal cells of the table would be lost entirely, but we have parallel data for all 75 ridings (i.e., electoral districts) in the province. The challenge of ecological inference, then, is to estimate Table 1’s missing values by observing how riding support for sovereignty varies with a group’s

share of the population. This is a risky venture, since an unfortunate pattern of aggregation could produce deceptive correlations between group density and voting behavior. But experts armed with both detailed contextual knowledge and adequate statistical techniques should be able to steer clear of naive estimates.

The problem is, until quite recently the techniques available for this purpose were inadequate to the task. Goodman’s ecological regression is the most common method of estimating how people voted. Although researchers have tried a host of refinements, most applications of Goodman’s method possess numerous well-known flaws, five of which are relevant here. The approach:

- Assumes that a group votes the same way in each riding. This is blatantly false. Two ridings, Beauce and Lac Saint Jean, consist almost entirely of francophones. Whereas the latter voted overwhelmingly for sovereignty (74.1%), only 43.7% of voters in Beauce did so. Clearly francophone preferences varied.

- Allows impossible vote predictions (without artificial corrections, at least). For example, our application of ecological regression predicts that -10% of linguistic minorities supported sovereignty, an absurdity that EI prevents.

- Produces only a single, provincial voting estimate for each group. Unlike with King’s approach, the researcher loses the rich geographical variation

present in government-collected statistics.

- Cannot handle the presence of “aggregation bias” in the data. If individual voting behavior is somehow related to a community’s linguistic composition – one source of aggregation bias – then estimates produced by ecological regression can be terribly wrong.⁴ Ecological regression offers little warning and no solution to this common problem, even though researchers often have good reason to expect its presence.

- Cannot handle when voters of one group are less consistent in their voting behavior than voters of another group (heteroskedasticity). Since francophone preferences varied widely from one riding to another, but linguistic minorities consistently opposed sovereignty, that statistical problem clearly occurs here.

The Francophone Vote: An Application of Gary King’s EI

Gary King’s solution to the ecological inference problem, called “EI,” was designed to avoid the many flaws of ecological regression. To estimate francophone support for sovereignty, we proceed in two steps – (1) Estimating the turnout for each group, and then (2) Estimating the vote among those who turned out. The data for the first stage are represented in Table 2, where we know how many people voted and know how many people are francophone, but

Table 1 -- Quebec's 1995 Sovereignty Vote

	For Sovereignty	Against Sovereignty	No Vote	Total Population
Francophones	?	?	?	5,763,366
Linguistic Minorities	?	?	?	1,132,604
All Linguistic Groups	2,302,510	1,336,210	3,257,250	6,895,970

Note: All marginals are real data obtained from elections officials in Quebec. Those not voting include minors; linguistic breakdowns for voting-age population were not available.

do not know how many francophones turned out. (This table is for Quebec as a whole, but again we possess parallel data from each of 75 ridings.)

EI begins by identifying the complete set of values that might fill a table's cells. The obvious first limit is that turnout rates for each linguistic group must fall between 0% and 100%, but the "method of bounds" allows even greater precision. For example, since 4.7 million people turned out, and only 1.1 million people in Quebec are linguistic minorities, no fewer than 3.6 million francophones could have voted across the province. That is, the "lower bound" on francophone turnout is about 61.5%. The upper bound is 81.1%. Even if no linguistic minorities voted, at most 4.7 million of the 5.8 million francophones could have turned out. Such bounds, when applied to each riding, do an even better job constraining provincial voting estimates – francophone turnout is bounded between 61.9% and 77.6% (analysis not shown).

Each possible turnout level for one group is paired with a unique turnout rate for the other. At the provincial level, for example, if exactly 4 million francophones voted then we know 676,454 linguistic minorities must have – no other number would produce the total turnout. The proportional turnout rates are similarly linear. If we graph francophone voting rates by other voting rates, then, each riding will be represented by a line segment, the set of all possible turnout combinations. Figure 1A presents a "tomography plot," King's name for the combined line segments of all 75 ridings. This plot summarizes all deterministic information contained in the election data; no assumptions were required to produce it. Vertical lines – that is, lines with very narrow bounds for francophones – correspond to ridings dominated by that group. The ridings contain so few linguistic minorities that we know quite precisely how francophones voted. Horizontal lines represent ridings with few francophones; we are unsure how many turned out because the riding doesn't contain enough of them to impact the totals

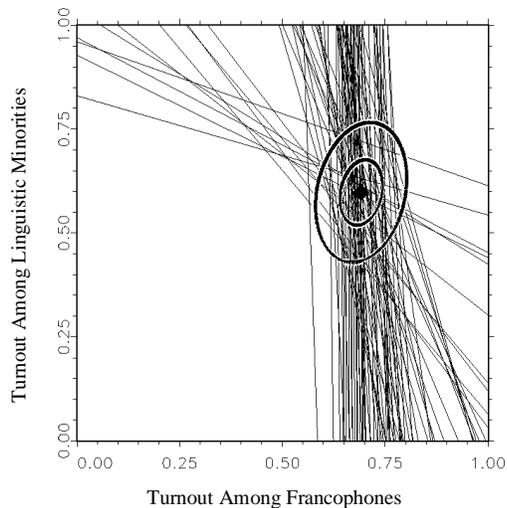
Table 2 -- Estimating the Turnout in 1995

	Voted	No Vote	Total Population
Francophones	?	?	5,763,366
Linguistic Minorities	?	?	1,132,604
All Linguistic Groups	4,676,454	2,219,516	6,895,970

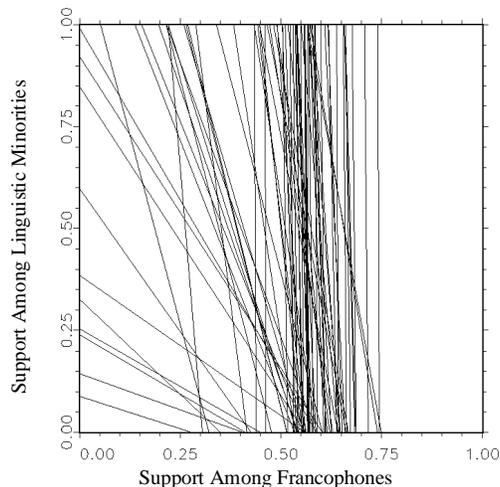
Note: All marginals are real data obtained from elections officials in Quebec. Those not voting include minors; linguistic breakdowns for voting-age population were not available.

Figure 1 - Tomography Plots for Turnout and Vote, Quebec 1995

(A) Turnout



(B) Vote



The ovals in Figure 1A represent contours for a bivariate normal probability distribution estimated by EI, and mark off the 80% and 95% confidence intervals much as contours do on an aerial map. The slanted lines cascading left on Figure 1B indicate aggregation bias in the voting data.

much. Angled lines portray ridings with a reasonable linguistic mix.

If we knew the exact turnout rates in Quebec's ridings, each would appear as a dot (rather than a segment) in the X-Y graph – but the aggregation process has prevented further narrowing of the options. Deriving more specific turnout estimates is impossible without making assumptions of some kind. The logical assumption, and the one EI makes, is that the real points tend to cluster wherever the lines converge. More formally, EI assumes a particular underlying pattern for the processes determining riding turnout, a roughly bell-shaped probability distribution called the “truncated bivariate normal.” EI estimates the most likely parameters for this bell-shaped curve, based upon where the tomography lines cluster, but then uses this estimated distribution to pick the highest-probability point on each line segment. In effect, the location of the other tomography lines determines the point estimate on any one.

Obviously this assumed pattern could be invalid in any one case. For example, if Quebec's ridings consisted of local tyrannies, turnout might be a direct function of who lived in each. Where francophones were a majority, voting by others would be suppressed almost completely. Where francophones were a minority, they would be disenfranchised. This underlying dynamic would not at all resemble a bell shape; turnout in each riding would approach the extremes. But any researcher with reasonable contextual knowledge knows whether such bizarre conditions prevail in the area under study. In Quebec, which lacks such idiosyncrasies, the multitude of minor factors adding up to voter turnout likely approach a bivariate normal distribution. The ovals in Figure 1A – which are contour lines, representing the estimated bivariate normal much as contours portray hills on an aerial map – probably capture the underlying variability governing turnout fairly well.

Once EI estimates turnout for each riding, we can proceed to the next step: estimating the vote of those who turned out. This procedure is almost identical

Table 3 -- Estimating the Vote for Sovereignty

	For Sovereignty	Against Sovereignty	Total Voters
Francophones	?	?	3,993,776 (25,166)
Linguistic Minorities	?	?	682,678 (25,166)
Total Vote	2,302,510	2,373,944	4,676,454

Note: The breakdown of voters by ethnic group represents estimates generated using Gary King's solution to the ecological inference problem, or EI. The standard error around those estimates is reported in parentheses.

to the previous stage. We know the distribution of votes, and we possess reliable estimates of turnout for each linguistic group (from the last stage), but we don't know the voting preferences of each group. EI fills in the missing cells of Table 3 the same way it dealt with Table 2 (although using “multiple imputation” to account for the additional uncertainty that comes from using estimated turnout rates).

EI obviously follows a complicated process. Using it to estimate francophone preferences may seem silly, the equivalent of building a cannon to kill a cockroach. Why not just assume that only francophones voted for sovereignty, since it held little appeal to others, and derive estimates for each riding based upon that assumption? In this case, the first stage of EI was more important than the second. Assuming an equal turnout rate for both francophones and others could result in grossly overestimating or underestimating francophone opposition to sovereignty, especially in heterogeneous ridings. Even if we have a strong theoretical prior that only francophones would support sovereignty – apparently an incorrect assumption, as it turns out – EI still helps determine how many *opposed* it. When the process is completed, we are left with relatively reliable estimates of francophone voting behavior in all 75 ridings. We turn briefly

to an analysis of those results in the next section.

Interpreting the Findings: The Effect of Context

Our estimates for the provincial sovereignty vote, found in Table 4, are fairly straightforward. We estimate that 55.6% of francophone voters and 11.9% of other voters supported sovereignty. The latter figure seems troublesome, because it is so high, but the large standard error on the estimate properly indicates our uncertainty.

Because we could get similar, and probably more reliable, figures from survey data, these estimates are not inherently valuable. However, we can compare them to estimates produced by ecological regression. Goodman's approach suggests that, while a reasonable 59.6% of francophones backed sovereignty, a negative number (-9.9%) of others did so. Furthermore, the standard errors deceptively imply great certainty. This guess is more than 3 standard deviations away from our EI estimate for the francophone vote, and the Goodman's voting estimate for linguistic minorities is supposedly two-and-a-half standard deviations away from zero, the nearest possible result!

The difference between these two methods is not merely a technical quibble; it contains real substantive meaning. Our

Table 4a -- Estimated Sovereignty Vote: 1995

	For Sovereignty	Against Sovereignty	No Vote	Total Population
Francophones	38.5%	30.8%	30.7% (0.44)	5,763,366 100%
Linguistic Minorities	7.2%	53.1%	39.7% (2.22)	1,132,604 100%
All Linguistic Groups	2,302,510	1,336,210	3,257,250	6,895,970

Table 4b -- Voter Sovereignty Preferences

	For Sovereignty	Against Sovereignty	Total Voters
Francophones	55.6% (3.4)	44.4%	3,993,776 100%
Linguistic Minorities	11.9% (19.6)	88.1%	682,678 100%
All Linguistic Groups	2,302,510	1,336,210	

Note: All marginals are real data obtained from elections officials in Quebec. All interior cells represent estimates generated using Gary King's solution to the ecological inference problem. Standard errors are reported in parentheses.

original interest in this data was to test whether francophone regionalism increased or decreased in heterogeneous ridings. In keeping with Group Conflict theory, francophones in mixed ridings might be the strongest supporters of sovereignty, because they are more likely to compete with Canada's dominant Anglo population for jobs or for control of their localities. On the other hand, francophones in mixed ridings may have adjusted to the stresses of living between two worlds already, and indeed may have economic or social ties to the Canadian majority. Those living in ethnic enclaves may feel besieged, without the mitigating effect of cross-cultural contact, and have less to lose from severing provincial ties with Ottawa.⁵

Naive ecological regression proves completely unhelpful at exploring those competing hypotheses, because it assumes constant francophone preferences across ridings. More importantly, though, it collapses precisely because francophone behavior is *not* constant across ridings. Goodman's method fits a line, as shown in Figure 2, to represent changes in the vote as a riding's linguistic mix changes. But the linear decrease in sovereignty support, as the francophone proportion drops, is much too strong for it to be capturing different ethnic preferences alone. Francophones are abandoning sovereignty in the mixed ridings, and ecological regression is falsely attributing the entire vote decline to the additional lin-

guistic minorities.

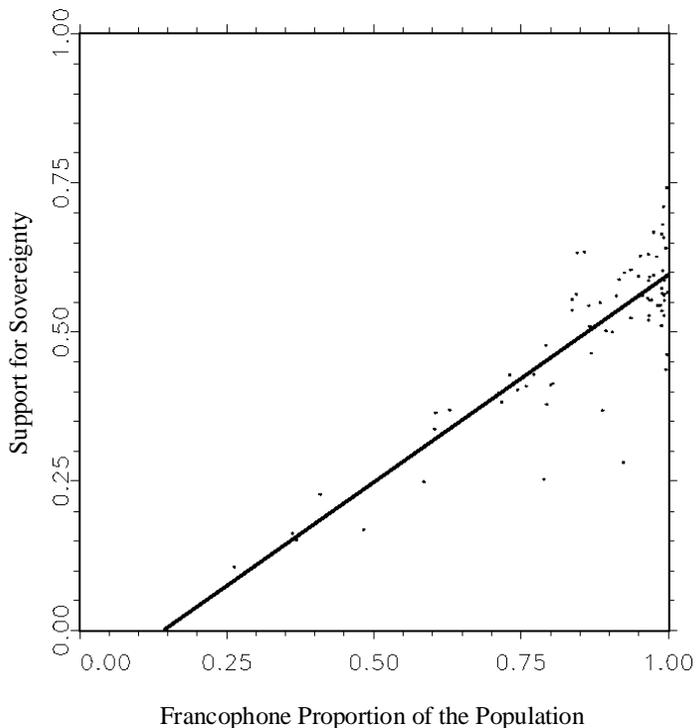
By contrast, even if we do not use a fancier version of EI that accounts for aggregation bias explicitly, King's method does not lead us so far astray. Figure 1B illustrates how, by applying the method of bounds to each riding, EI remains robust in the face of aggregation bias. The vertical lines, which cluster to the right of the plot, indicate strong sovereignty support in most francophone-dominated ridings. The angled lines, though, which represent mixed ridings, cascade to the left. Even the strongest possible francophone support for sovereignty in these ridings *must* be lower than that found among homogeneous constituencies, because the angled lines do not intersect with most of the vertical ones. By incorporating this known information through the method of bounds, therefore, EI captures some of the dynamic that undermined ecological regression.

Figure 3 shows the beneficial effect of taking these bounds into account. The Y axis of this graph represents our EI estimates of francophone support for sovereignty. The X axis represents the ethnic makeup of each riding. The line fit through these points summarizes the linear trend in our estimates. Whereas Goodman's model necessarily imposes an unrealistically straight line here, since it only produces one estimate for the entire province, EI picks up a substantively interesting phenomenon: that francophones with the most interethnic contact were least likely to support sovereignty.

EI and the Future of Comparative Research

Aggregate data collected by governments are a rich, and usually cheap, source of information. Unfortunately, individual-level detail is lost from this source. Sometimes researchers finesse the loss of information, choosing to study actions of provinces rather than people, but not all theories of political behavior are so tractable. The Group Conflict theory tested in the last section is one example. Since the theory describes individual responses to the community context, changing the focus of study is not credible. Leaving voting data at the ag-

Figure 2 - Aggregation Bias and Negative Estimates



The regression line represents results of Goodman's ecological regression performed on riding-level data from the 1995 Quebec sovereignty vote. Because the line does not reach the left vertical border within the unit square, estimated support for sovereignty among linguistic minorities is impossibly negative, a result of aggregation bias.

gregate level simply does not capture the relevant political behavior.

Some researchers in comparative politics, faced with such difficulties, already have begun taking advantage of the new leverage EI provides. For example Daniel Posner, a Harvard doctoral candidate studying ethnic politics in the developing world, has formulated a theory connecting the political salience of ethnic identity to changes in Zambia's electoral system. Whereas the one-party system of the 1970's and 1980's encouraged patronage-oriented Zambians to vote their tribal loyalties, the multi-party system of the 1990's promotes identification with wider-ranging linguistic groupings. Numerous policy implications for managing ethnic strife obviously spin off of the hypothesis, if empirical evidence supports it.

Posner has collected constituency-level voting returns from Zambia's 1983 and 1988 one-party elections and 1991 and 1996 multi-party elections, as well as identified the tribal background of

every parliamentary candidate in each election. He also has parallel 1990 census figures for tribal membership, aggregated up from wards. But the observable implications of the theory operate at the individual level, not at the level of entire constituencies.

EI has allowed a reasonable test of Posner's hypothesis using this data. For each election, he estimates the extent to which a constituency's dominant tribe voted for candidates who share their affiliation, and compares that rate to an estimate of how much other tribes supported the same set of candidates. The greater salience of tribal ties, the less crossover voting should appear. Posner's preliminary analysis shows exactly what he anticipated – tribal polarization in the one-party elections, less polarization in the multi-party elections. The use of EI paid off.

Gary King's solution to the ecological inference problem is not perfect. For example, it is computationally intensive. Complex versions of EI applied to ex-

tensive data can tie up slower computers for an entire day. It also requires a greater degree of methodological sophistication from the user than did less-satisfying approaches to ecological analysis. Often those older methods derive perfectly adequate estimates of group behavior, although perhaps more through good luck than any virtue in the method. Finally, the development is a new one; certainly it will improve as additional methodologists turn their attention to King's work.

Nevertheless, using EI for ecological analysis should soon become the norm in comparative politics. Relative to the alternates that are currently available, the drawbacks to King's method are all practical ones. On methodological grounds, EI leaves its predecessors far behind. King also has worked hard to make the approach accessible, providing code from a statistical software package called GAUSS that implements the method, as well as a stripped-down version for PC users who lack GAUSS (the URL is <http://gking.harvard.edu>). But above all, King's EI should spread because it allows the study of comparative political behavior to focus where it often belongs – on the behavior of individuals within their political system.

¹The authors are indebted to Gary King for his mentorship, and especially his instruction in ecological inference. We present his innovations here not as objective observers but as acolytes, aiding the cause only by making the intuitions more accessible. All errors and shortcomings in the presentation are ours.

²Commercial surveys in the U.S. usually only sample a handful of people per county. See: Voss, D. Stephen, Andrew Gelman and Gary King. 1995. "Preelection Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992." *Public Opinion Quarterly* 59: 98-132.

³ King introduces the method in: King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior From Aggregate Data*. Princeton, NJ:

Princeton University Press. A formal comparison of EI's performance to that of rival methods, in a case with very good electoral data, appears in: Palmquist, Bradley L. and D. Stephen Voss. 1996. "Racial Polarization and Turnout in Louisiana: New Insights from Aggregate-Data Analysis." Paper presented at the 54th Midwest Political Science Association meeting, Chicago, IL, April 18-20. Copy available upon request.

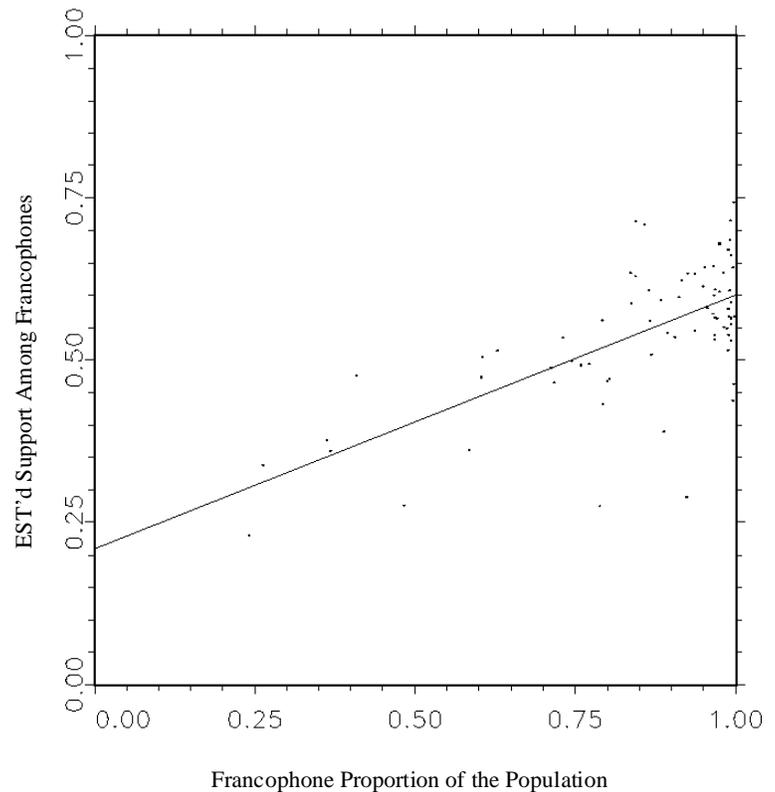
⁴The problems with aggregate-data analysis were popularized in: Robinson, W.S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15:351-57. Although not widely available, one source with a nicely intuitive treatment of ecological inference's dangers is: Palmquist, Brad. 1993. "Ecological Inference, Aggregate Data Analysis of U.S. Elections, and the Socialist Party of America." Unpublished Ph.D. dissertation, University of California at Berkeley.

⁵A brief introduction to the Group Threat theory appears in: Giles, Micheal and Arthur Evans. 1986. "The Power Approach to Intergroup Hostility." *Journal of Conflict Resolution* 30:469-86. Some methodological issues related to ecological inference that spring up when testing the theory are discussed in: Voss, D. Stephen. 1996. "Beyond Racial Threat: Failure of an Old Hypothesis in the New South." *Journal of Politics* 58:1156-70. For a discus-

sion of the Contact Hypothesis as it applies to Canada, see: Berry, J.W. 1984. "Cultural Relations in Plural Societies: Alternatives to Segregation and Their Sociopsychological Implica-

tions." In *Groups in Contact: The Psychology of Desegregation*, ed. Norman Miller and Marilyn B. Brewer. Orlando, FL: University Press of America, pp 11-27.

Figure 3 - EI's Robustness in the Face of Bias



Gary King's solution to the ecological inference problem is robust in the presence of aggregation bias, even when no explicit corrections are made for that bias. The method of bounds permits estimated sovereignty support among francophones to vary as the linguistic makeup of a riding changes.

Datasets & Archives

The International Social Survey Program

Tom W. Smith

*National Opinion Research Center
University of Chicago*

The International Social Survey Program (ISSP) is a continuing, annual program of crossnational collaboration. It brings together pre-existing, social science projects and coordinates research

goals, thereby adding a crossnational perspective to the individual, national studies.

ISSP evolved from a bilateral collaboration between the Allgemeinen Bevölkerungsumfragen der Sozialwissenschaften (ALLBUS) of the Zentrum fuer Umfragen, Methoden und Analysen (ZUMA) in Mannheim, West Germany, and the General Social Survey (GSS) of the National Opinion Research Center (NORC) at the Univer-

sity of Chicago. Both the ALLBUS and the GSS are replicating, time series studies. The ALLBUS has been conducted biennially since 1980 and the GSS annually (except for 1979 and 1981) since 1972. In 1982 ZUMA and the NORC devoted a small segment of the ALLBUS and GSS to a common set of questions on job values, important areas of life, abortion and feminism. (A merged dataset is available from the Interuniversity Consortium for Political and So-

cial Research (ICPSR) at the University of Michigan.) Again in 1984 collaboration was carried out, this time on class differences, equality and the welfare state.

Meanwhile, in late 1983, Social and Community Planning Research (SCPR) in London, which was starting a social indicators series called the British Social Attitudes Survey (BSA) similar to the ALLBUS and GSS, secured funds from the Nuffield Foundation to hold meetings to further international collaboration. Representatives from ZUMA, NORC, SCPR and the Research School of Social Sciences at the Australian National University organized ISSP in 1984 and agreed to (1) jointly develop topical modules dealing with important areas of social science, (2) field the modules as a fifteen minute supplement to the regular national surveys (or a special survey if necessary), (3) include an extensive common core of background variables and (4) make the data available to the social science community as soon as possible.

Each research organization funds all of its own costs. There are no central funds. The merging of the data into a crossnational dataset is performed by the Zentralarchiv fuer Empirische Sozialforschung at the University of Cologne.

Since 1984, ISSP has grown to 29 nations, the founding four—Germany, the United States, Great Britain, and Australia—plus Austria, Italy, Ireland, Hungary, the Netherlands, Israel, Norway, the Philippines, New Zealand, Russia, Japan, Bulgaria, Canada, the Czech Republic, Slovenia, Poland, Sweden, Spain, Cyprus, France, Portugal, Slovakia, Latvia, Chile and Bangladesh. In addition, East Germany was added to the German sample upon reunification. The affiliated organizations are listed in Table 1. Other nations have replicated particular modules without being ISSP members (Poland, in 1987, and Switzerland, in 1987 and 1993).

The annual topics for or ISSP are developed over several years by a subcommittee and pretested in various coun-

tries. The annual plenary meeting of ISSP then adopts the final questionnaire. The ISSP researchers especially concentrate on developing the questions that are (1) meaningful and relevant to all countries and (2) can be expressed in an equivalent manner in all relevant languages. The questionnaire is originally drafted in British English and then translated to other languages using standard back translation procedures.

The themes covered in the ISSP module and the nations collecting data are listed in Table 1. The first theme on the role of government covered attitudes towards (a) civil liberties, (b) education and parenting, (c) welfare and social equality and (d) the economy. The second theme was on social networks and support systems. It contained detailed behavioral reports on contacts with various friends and relatives and then a series of questions about where one would turn for help when faced with various situations such as financial need, minor illness, career advice and emotional distress. The third module, on social equality, concerned beliefs about what factors affect one's chances for social mobility (e.g. parental status, education, contacts, race, etc.), explanations for inequality, assessments of social conflicts and related questions. It also asked people to estimate the average earnings of various occupations (e.g. farm laborer and doctor) and what the average earnings of these occupations should be. The fourth module covered the impact on the family of the changing labor force participation of women. It included attitudes on marriage and cohabitation, divorce, children and child care and special demographics on labor force status, child care and earnings of husband and wife. The fifth module on orientations towards work dealt with motivations to work, desired characteristics of a job, problems relating to unemployment, satisfaction with one's own job (if employed) and working conditions (if employed). The sixth module in 1990 repeats the role of government theme. By replicating substantial parts of earlier modules, ISSP will not only have a crossnational per-

spective, but also an over-time perspective. We will not only be able to compare nations and test whether similar social science models operate across societies, but also be able to see if there are similar international trends and whether parallel models of social change operate across nations. The seventh module covers the impact of religious beliefs and behaviors on social, political, and moral attitudes. It includes questions on religious upbringing, current religious activities, traditional Christian beliefs and existential beliefs. The non-religious items concern such topics as personal morality, sex roles, crime and punishment and abortion. The eighth module in 1992 replicates and extends the 1987 social equality module. The ninth module in 1993 is on the environment. It includes an environmental knowledge scale along with attitudinal and behavioral measures. The tenth module in 1994 repeats the 1988 module on women, work and the family. It also adds items on household division of labor, sexual harassment and public policy regarding the family. The eleventh module in 1995 was on national identity. It assesses nationalism and patriotism, localism and globalism, and diversity and immigration. 1996 will be the second replication of the role of government, 1997 will be the first replication of the 1989 module on work orientations, 1998 the first replication of the 1991 religion module, 1999 the second replication of the 1987 and 1992 social inequality modules, and 2000 the first replication of the 1993 environment module.

ISSP marks several new departures in the area of crossnational research. First, the collaboration between organizations is not special or intermittent, but routine and continual. Second, while necessarily more circumscribed than collaboration dedicated solely to crossnational research on a single topic, ISSP makes crossnational research a basic part of the national research agenda of each participating country. Third, by combining a crosstime with a crossnational perspective, two powerful research designs are being used to study societal processes.

Good Reads

David D. Laitin
University of Chicago

Data from the first ten modules on role of government, social networks and support systems, social equality, the family, work orientation, role of government II, religion, social equality II, the environment and the family II are presently available from the Zentralarchiv and various national archives such as Essex in Britain and ICPSR in the United States. The 1995 national identity module will be available shortly and the other modules will be released periodically as soon as the data can be processed.

Publications based on the ISSP are listed in a bibliography available from the ISSP Secretariat (see below).

Three collections of ISSP research are (1) Roger Jowell, Sharon Witherspoon and Lindsay Brook, eds., *British Social Attitudes: Special International Report*. (Aldershot: Gower, 1989); (2) J.W. Becker, James A. Davis, Peter Ester and Peter P. Mohler, eds., *Attitudes to Inequality and the Role of Government*. (Rijswijk, The Netherlands: Sociaal en Cultureel Planbureau, 1990); and (3) Roger Jowell, Lindsay Brook and Lizanne Dowds, eds., *International Social Attitudes: The 10th BSA Report*. (Aldershot: Dartmouth Publishing, 1993).

For further details contact the ISSP Secretariat, Tom W. Smith, NORC, 1155 East 60th St., Chicago, IL, 60637. Phone: 773/256-6288. Fax: 773/753-7866. Email: smitht@norcmail.uchicago.edu.

Good Reads

With this issue, the Newsletter inaugurates a new column, "Good Reads." Officers of the Organized Section will from time to time contribute discussions of works that have significantly inspired and influenced their own research. These columns are not meant as formal reviews, but instead to alert readers to important and useful ideas and findings.

It might surprise avid readers of this section newsletter – a newsletter in which I have been part of a governing junta that has sought to infuse microanalytic models into the habitus of comparative politics – that I co-edit (with George Steinmetz, a sociologist at the University of Michigan) a book series published by Cornell University Press and called "The Wilder House Series in Politics, History and Culture." At the risk of self-promotion, I want to highlight a few of the books in this series as "good reads."

What distinguishes the Wilder House series is its interdisciplinary exploration of the foundations of nations and states. Methodologically, unlike most studies in the microanalytic tradition, Wilder House books reflect intensive exposure to a case or a few cases. Thus there is far more narrative and descriptive inference and far less concern for parsimony and causal inference. Substantively, series coherence is achieved by bringing authors to Wilder House (at the University of Chicago), usually before the final rewrite of an accepted manuscript, and having them listen to the discussion of the editorial board and interns (made up of faculty and advanced graduate students in political science, sociology, history and anthropology) about the book and its principal arguments. In the final rewrite, authors tend to incorporate into their manuscripts more explicitly than would have happened otherwise their ideas on how culture and historical contexts shape the contours of nations and states.

The books in the series have broad relevance to students of comparative politics. While there is no way I can review all the books in the series, I'll try here to give a taste of what we have produced. Several of the books focus upon the making (and unmaking) of states. James Given's *State and Society in Medieval Europe* (1990) shows rather counter-intuitively that as states expand, the decision as to whether to institute direct or indirect rule has more to do with the social structure of the captured territory than the institutional capacity of the expanding state. Given is a medievalist historian and the book is therefore compelling in primary research. Yet his comparison of Wales and Languedoc has the aura of a natural experiment that is sure to impress political sci-

tists. Victor Magagna's *Communities of Grain* (1991) reveals a community (rather than class) basis of revolutionary action and the concomitant undermining of state authority. The historical and comparative range of the materials is matched by a clear theoretical focus on how communities act collectively in violent assaults against state power. Ian Lustick's *Unsettled States, Disputed Lands* (1993) jars our sense of the normal by showing the contingency of state territorial boundaries and the possibilities of hegemonic projects to undermine what had once been conceived of as "natural." The case studies of Ireland, Algeria and the West Bank are rich in historical detail yet tied to a sophisticated threshold model of institutionalization. Karen Barkey's award-winning *Bandits and Bureaucrats* (1994) relies on primary sources from the Ottoman Empire in order to demonstrate that contra Tilly, war is not a necessary precondition for state construction; focus on nonwestern cases, the book demonstrates, allows for a broader set of paths toward modern centralized rule. These books, focusing on the making (and unmaking) of states – all with an historical focus – make not only for exemplary social science, but also, due to the employment of narrative, "good reads" as well.

A major focus for many of the books in the Wilder House series is on the formation of nations and national cultures, especially in regard to language and religion. In fact, the flagship volume for the series is a magnificent set of essays by Benedict Anderson *Language and Power* (1990), which is prefaced with an autobiographical essay that tellingly reveals the intellectual divide between the author and his brother, Perry, on issues of culture and class and of state and nation. William Miles' *Hausaland Divided* (1994), by examining microscopically two communities under different colonial regimes (a French regime in Niger; a British regime in Nigeria) shows how forms of colonial rule shape institutions and re-shape national cultures. A complementary book to Miles' is Juan Díez Medrano's *Divided Nations* (1995) where the author traces the divergent social bases of Basque and Catalan nationalism within the Spanish state. Janet Hart in her *New Voices in the Nation* (1996) narrates – most often in the telling

voices of her subjects –the story of how women who participated in the Greek resistance got incorporated into and were ultimately marginalized by the Greek nation.

Herman Lebovics has perhaps written the quintessential Wilder House book, *True France* (1992). In it, he demonstrates how third republic France created rules of both national inclusion and exclusion. On the one hand, it was able to make “Frenchmen into peasants” by creating an ideologically conservative model of the French nation. On the other hand, third republic anthropology helped naturalize the quasi-exclusion the colonized nations such as the Vietnamese. In a similar study, Mabel Berezin’s *Making the Fascist Self* (1997) explores the creation of a fascist identity in Italy, and depicts the public rituals and the calendar of festivities that worked to create a “community of feeling.” Through an examination of letters from soldiers on the front, Berezin was able to show as well the limits of the fascist project. Stathis Kalyvas’ award winning *The Rise of Christian Democracy in Europe* (1996) explains how religious parties emerged in 19th century Europe despite the reluctance of the Church to get involved in politics, and how those parties played an inadvertent role in promoting European secularism. Frederic Schaffer’s *Democracy in Translation* (1998) ethnographic research in Senegal shows how institutions such as “democracy” have different meanings depending on whether discourse about it is in French or Wolof. If international or domestic pressure requires deepening of democracy, what would be deepened, he shows, is quite different depending on whether the French word or its Wolof quasi-equivalent is used. These books on the nation and national cultures expand our case material, show how national and cultural ideas emerged, and how they fared as hegemonic projects.

In press is an edited volume by George Steinmetz, *State\Culture*, which in many ways sums up the first decade of Wilder House publications and sets an agenda for the future. What has made this series a joy to edit, as I hope I’ve made clear, is that the books are not only contributions to comparative politics and historical sociology, but they are fun to read as well.

Book Reviews

Traditional Politics and Regime Change in Brazil

Frances Hagopian
Cambridge University Press
Cambridge, 1996

Reviewed by Ivani Vassoler
University of Maryland
bem@wam.umd.edu

If we accept the assumption that times of regime change constitute one of the best moments for the creation of a new political system by dismantling the pattern of traditional politics, Brazil is an exemplary case of missed opportunities. In a thirty-year period during which its regime changed twice – to military rule and later to civilian government – Brazil’s political structure did not change as scholars might have expected: rather political continuity with the dominance of traditional elites in politics is what characterizes the country today. This is the central thesis advanced by Frances Hagopian in *Traditional Politics and Regime Change in Brazil*. To explain why regime change did not lead to political change in Brazil, Hagopian examines the main developments of the country’s military rule (1964-1985) and the democratization process that took place after twenty-one years of dictatorship. Both regimes, she argues, were not able to promote political change since the traditional elites not only have survived the transformations, but they also have played an important role in them.

Hagopian’s claim that despite regime change the old political structure remains the same is sustained through an examination in depth of the politics of Minas Gerais, a state located in the center of the Brazilian vast territory. Although focusing her analysis on only one of the country’s 26 states, Hagopian fully captures the essence of Brazil’s political life, which is dominated by the traditional elites. With a powerful political class and a conservative society, Minas Gerais is a sort of cradle of clientelism, regionalism and personalism, the three features the author identifies as the basis of the traditional elites. These three characteristics, which eventually explain the political continuity argument, also can be found in the other Brazil’s states.

Besides enhancing the understanding of Brazilian politics, Hagopian’s book contributes to the field of comparative politics by contrasting the largest Latin American country’s political developments with those of Uruguay, Argentina and Chile, which also experienced a period of authoritarian rule followed by a process of democratization. Yet Hagopian’s separated analyses – one focused on Brazil and the other on its neighbors – come with some controversial premises and findings.

Beginning with the case of Brazil the author correctly stresses the dominant role of the traditional elites in the country’s politics. Yet, the assumption that the 1964 regime change – from democracy to dictatorship – was a period in which political change could also occur is largely disputed by several scholars. For Hagopian, the dictators failed in their attempts to revamp the political system because they could not eradicate the old elites from power. This is absolute true. What it is a mistake is to suppose that the five generals that occupied the presidency until 1985 had ever in mind to wage a war against traditional politics. In fact, in the way that the successive authoritarian governments evolved, one is more inclined to think that the military was more willing to preserve the traditional political elite than to eliminate it. Through clientelism – the allocation of state resources in exchange for votes – the oligarchy helped the military to survive – for sometime – and to sustain the state capitalist model. It is a classic case of one sector reinforcing another, since as Hagopian remarks, the regime relied on the traditional elites to remain in power; and by limiting political competition the regime also “helped traditional political elites to retain their power and position”.

In some sense the absence of political change during the authoritarian rule was a blessing because the military’s inability to transform the political structures eventually contributed to the regime demise; more specifically by protecting the traditional elites, the authoritarian regime undermined itself. The withdrawal of the military from power is a reason for celebration. Less exciting are the consequences of the Brazilian traditional elites’ participation in the construction of the new regime.

The traditional elites’ role in the transi-

tion to democracy should not be a surprise: enhanced by the support of the old rulers, the elite emerged stronger in the democratization process and conducted it, hence obscuring the prospects for political change. Thus following Hagopian's analysis, for a second time in three decades another opportunity was missed in Brazil for a radical transformation in its political system. As occurred with the previous regime, the new one failed to reshape the state-society relations and to alter the patterns of competition and representation, through a modern, plural and democratic politics.

The role of the traditional elites in Brazil's democratization process is undeniable and Hagopian correctly states that the current president Fernando Henrique Cardoso was elected with the support of the traditional politicians. What again remains doubtful, however, is the hypothesis that those involved in the transition to democracy had political change in mind. Are the traditional politicians willing to transform the country's political structure? As the Brazilian democratic reforms are being conducted mainly by the traditional elite, it is more likely to assume that the masters of the process are not plotting against themselves. This consideration, however, does not dismiss Hagopian's main thesis pointing out the existence of political continuity in Brazil. Indeed, after surviving two regime changes, the Brazilian traditional elites are alive and well. In this sense, the author makes an important contribution to the understanding of the relationship between regime change and political change, showing through the Brazil's example, that the latter is not the immediate consequence of the former.

Regarding the experience of the other Southern Cone bureaucratic-authoritarian countries and their subsequent democratic transitions, Hagopian arrives at the conclusion that those governments were more successful than Brazil in producing political change. Unfortunately the analysis does not show strong evidence to support this claim. In fact, while political continuity is remarkable and visible in Brazil, it is not so clear that in Uruguay, Argentina and Chile a real transformation is occurring in the state-society relations. Certainly it is wrong to dispute Hagopian's statement that in Brazil the control of the traditional elites and the exaggerated state clientelism are diminishing the people's faith in political institutions. Yet, is such state of affairs too different from that prevalent in the Brazil's neighbors? Undoubtedly there are many flaws in the Brazil-

ian democracy, among them the fact that the new regime is constrained by the power of the traditional politicians. But nobody should overlook the fact that the other emergent democracies in the region also face serious obstacles to their consolidation.

After her exhaustive analysis, it is understandable Hagopian's concern with the future of the Brazilian democratization process in face of the traditional elites' dominance. The political continuity justifies the pessimism shared by many – including this writer – regarding Brazil's consolidation of democracy. The untouchable power of the elites constitutes certainly a great obstacle to real democratic development not only in Brazil, but in several Latin American countries as well.

Heroic Defeats: The Politics of Job Loss

Miriam A. Golden
Cambridge University Press
Cambridge, 1997

*Reviewed by M. Victoria Murillo
Harvard University
mmurillo@fas.harvard.edu*

Golden develops a game-theoretical model to explain the outbreak of strikes over large-scale job losses. These strikes are costly for both unions and firms, and invariably end in heroic defeats for the unions. At first glance, these strikes appear irrational: why would labor enter into what is clearly a losing proposition? Golden molds the interaction between unions and firms to demonstrate the 'hidden' rationality of strikes against job losses. She then tests her model using comparative case studies.

Golden focuses on the goals of union leaders rather than the interests of workers to illuminate the logic of union survival motivating the strikes. Strikes, as a form of collective action, require organization by a union. Unions are not merely the agent of workers but have their own institutional goals, including the primary goal of organizational survival. Union leaders organize strikes over redundancies when job losses target union activists, and threaten the survival of the unions themselves. Union leaders persuade workers to go on strikes even though the strikes have little hope of defending job losses because they want to defend the existence of unions. The rank-and-file members follow union leaders because they lack impartial information about

their opportunities to win a strike or because they are divided on their perceptions of redundancies.

Golden's model is based on three assumptions: firms prefer to dismiss union activists to recover management discretion; strikes are costly both for firms and unions; and union leaders only go on strike to defend activists who influence union membership (which determines union revenue and leaders' potential income) (Golden, 18). From the preferences of each party, she derives two pairs of testable propositions. Strikes over job losses are more likely to occur either when the firm does not know the union threshold for tolerating the firing of activists, or when both parties prefer to strike because a third party subsidizes the cost of the conflict. In contrast, strikes over job loss will never occur if the firm is prevented from targeting union activists by an external rule (e.g. seniority), or if the union has a reputation of being strike-prone, which prevents the firm from targeting activists to avoid the costs of a strike.

Golden employs the comparative method to test the propositions inferred from her model. First, she compares two similar cases with different outcomes in the automobile sector after the second oil crisis of 1979: the strike against Fiat lay-offs in Italy and the non-conflictive downsizing of British Leyland in Britain. Then, she contrasts different cases with the same outcome: the strikes against redundancies in the British coal mines of Yorkshire in 1984-85, and the strikes against downsizing in the Japanese Miike coal mine in 1959-60. By analyzing these cases, she shows that union leaders responded similarly to common incentives despite contextual differences in political systems, culture, and industrial relations systems.

Golden carefully describes the historic events of each case study to present the evidence necessary to test her hypotheses. The first comparison considers the automobile sector. Fiat and British Leyland faced similarly fragmented union structures when reducing their work force. Yet, only the former experienced a conflict. Golden explains this different outcome by pointing to British Leyland's use of inverse seniority for shedding workers. This rule protected union activists by reducing management's discretion to target them. In contrast, Fiat inadvertently targeted too many activists, which surpassed the union tolerance threshold and triggered the strike.

The second comparison deals with the coal industry. Whereas the Japanese coal industry had been conflict-prone, its British counterpart had traditionally been peaceful. Both countries also had very different industrial relations systems. Despite these differences, job redundancies provoked a similar response in the two cases: bitter and long strikes in both the Japanese Mitsui mine of Miike in 1959-60, and the British mines of Yorkshire in 1984-85. In both cases, the targeting of activists was accompanied by support of national organizations which subsidized the conflict for firms.

The Japanese national business association Nikkerei subsidized the Mitsui firm during the Miike strike by providing financial help while competitors respected its markets. Employers supported the firm because coal was the only industry where a national trade union, Tanro, played a role in collective bargaining against the system of enterprise unionism preferred by business. In Britain, the new Thatcher government was prepared to back the British National Coal Board against a strike because it resented the National Union of Mineworkers' (NUM) hostility against the governments' plans to reorganize the coal industry and the national industrial relations system. Although the unions involved defended their targeted activists, the national labor organizations had broader goals. The Japanese confederation, Soyho, and the national union, Tanro, supported the Miike strike in an attempt to redefine collective bargaining patterns at the national level. The British NUM wanted to bring down the new Tory government and promote the election of a more sympathetic Labour government which would preserve a decaying industry and the strength of the NUM.

Golden concludes that firms take advantage of large job reduction to target union activists and restore management discretion. While strikes over job loss cannot prevent redundancies, they can defend the union as an institution. In addition, when national actors support specific strikes, the consequences of the conflict are broader. This is shown in the reinforcement of enterprise unionism in Japan and the weakening of labor unions in Britain. An extension of her argument also explains the predominance of concession bargaining over redundancies in the U.S. where seniority rules prevented firms from targeting union activists and reduced conflicts over large-scale job losses.

Although Golden's analysis of the union as an organization explains the rationality of strikes over job losses by focusing on the institutional goals of union leaders and the imperfection of their agency in relation to workers, some of her assumptions would benefit from further empirical support. For instance, she assumes that the targeting of activists threatens union leaders. However, she also describes how union leaders' accepted the firing of 'troublesome' activist, which in turn provoked the Fiat management to miscalculated the union tolerance threshold (Golden, 74). That is, she does not analyze the relationship between union activists and union leaders, although this relationship can strongly affect leader preferences. Moreover, she argues that rank-and-file misinformation and division provoked workers to follow leaders into heroic defeats. Yet, even if union leaders are imperfect agents, they are still agents of workers and are accountable to them to some extent, since workers may practices exit or voice. Although Golden states that the interests of the rank-and-file constrain, but do not

determine, union action (Golden, 27), she does not provide information on what those constraints on union action are and how they affect the strategies of union leaders. Finally, while rules over redundancy (either seniority or work sharing) are easy to measure ex-ante to evaluate their impact on protecting activists, the tolerance threshold of the union is not so easy to measure. Golden only measures it ex-post and infers its impact on union action, by the common incentives it creates for union leaders in different contexts. However, since the only way to see that the firm reached the threshold is the strike itself, it is hard to falsify this proposition.

In short, Golden offers an innovative view of the rationality of strikes over job losses by combining deductive inference from a game-theoretical model with comparative case studies to provide the evidence used to test her model. Her focus on unions as organizations and her method of analysis provide a powerful account of union leaders' actions within different contexts, and of the consequences of industrial conflicts for national industrial relations systems. Furthermore, her innovative approach is a very useful way of combining different methodological approaches to advance the understanding of comparative political economy. Golden shows, in a rigorous and empirical fashion, that theoretical concepts can both travel and grasp the historical specificity of the cases. Although her model cannot account for the internal dynamics of unions without eroding the parsimony of her explanation, her conceptualization of the union tolerance threshold may be refined with empirical indicators to make its measurement clearer in order to strengthen its explanatory power.