

COMPARING HISTORICAL SEQUENCES—A POWERFUL TOOL FOR CAUSAL ANALYSIS

A REPLY TO JOHN GOLDTHORPE'S "CURRENT
ISSUES IN COMPARATIVE MACROSOCIOLOGY"

Dietrich Rueschemeyer and John D. Stephens

ABSTRACT

John Goldthorpe is right when he identifies three major methodological problems of macrosociology—problems due to the small number of cases, problems due to the lack of full independence of the cases, and problems that arise when we move from association to causal explanation. He is wrong when he claims that comparative historical research is more plagued by these problems than cross-national quantitative research. In fact, while we advocate the integration of comparative historical and cross-national statistical analysis, we claim that comparative historical research has particular advantages in dealing with the problems identified because it focuses on historical sequence and because it can take account of varied historical contexts.

Comparative Social Research, Volume 16, pages 55-72.
Copyright © 1997 by JAI Press Inc.
All rights of reproduction in any form reserved.
ISBN: 0-7623-0250-X

John Goldthorpe discusses three central issues in comparative macro-sociology—the problem that typically only a small number of cases are available to disentangle complex causal patterns; “Galton’s problem” or the question whether the cases studied are independent of each other; and the “black box” problem or the question of how to move from findings of association to causal explanation. Goldthorpe is right to consider these serious problems. He is also right to see them as problems faced by both cross-national quantitative and comparative historical research. He is wrong, however, when he claims that comparative historical research has no advantage in dealing with these issues and when he in fact seeks to turn their discussion into an argument for the superiority of the cross-national statistical approach. Both modes of analysis have their own peculiar advantages as well as difficulties.¹ Nevertheless, we claim, however, that comparative historical research has a distinctive and critical advantage in macro-sociology because it focuses on historical sequence and because it allows one to take account of varied historical contexts. When in addition is joined to cross-national statistical research results and extended to as many cases as possible, we consider comparative historical analysis the method of choice for the study of macro-social phenomena.

Before we discuss the specific problems in greater detail, we offer a few comments on the place of historical research in macro-social analysis and on the issue of holism. We then briefly introduce the study in which we have recently collaborated and which we will use repeatedly as an instance of the research procedure we consider most appropriate.

ON THE NEED FOR HISTORICAL ANALYSIS

Navigare necesse est—the old assertion that sailing on the high seas, however risky, was necessary nevertheless can be transferred to the role of historical research in systematic social analysis. Historical research, too, is risky and fraught with problems: Historical evidence is drastically and irremediably incomplete; it is often of dubious validity; and it tends to be biased—favoring the victors rather than the losers, the lasting developments rather than the historical dead ends and detours, the rich and the educated rather than the poor and illiterate, and so forth. And yet for very cogent reasons historical research is not only useful, it is required for systematic social analysis, especially for macro-social research.

Why is historical research necessary? First, there is the fact that many social patterns, once formed, tend to persist. Many present-day conditions seem to have their roots in constellations several generations or even centuries old. Seymour Martin Lipset and Stein Rokkan came to this conclusion in their well-known analysis of voter alignments: “We simply cannot make sense of variations in current alignments without detailed data on differences in the

sequence of party formation and in the character of the alternatives presented to the electorate before and after the extension of the suffrage" Lipset and Rokkan (1967, p. 2). Wherever it may be necessary to reach in this way back into history in order to understand the present, all "presentist" explanations—explanations which consider only factors in the most immediate past—become profoundly suspect.

A second argument is closely related. Many phenomena of interest—and especially macro-social phenomena—are shaped by constellations of factors rather than just one in isolation. That means often that the sequencing of major causal conditions matters. Thus, it has been argued that class formation took a different path depending on the relative timing of industrialization and the extension of the suffrage (Katznelson and Zolberg 1986) or that the character of state organization in Europe depended profoundly on whether the rationalization of rule preceded the rise of rationalized capitalist production and exchange or vice versa (Nettl 1968).

True, we do not know enough about either historical persistence and its limits or the effects of different sequence patterns to use these as *explanatory* principles; but we do know enough to consider them *heuristic* principles of the first importance, especially in macro-analyses. As such, they make historical research indispensable.

Yet, the most important argument is much more elementary. Causation is a matter of sequence. One needs diachronic evidence, evidence about historical sequences, to explore and to test ideas about causation directly. This remains true even if it is also true that simple historical narrative is not the same as causal explanation. *Post hoc* does not any more translate directly into *propter hoc* than correlation demonstrates causation. Explanation can indeed not proceed without analytic hypotheses; but causal explanation ordinarily needs hypotheses about sequence which can best be tested against evidence about sequences. And the first two arguments advanced—about historical persistence and the relative timing of causal conditions—suggest that it is advisable to look for sequences beyond the immediate past.

Goldthorpe's argument rests on a radical separation of a narrative account of "what happened," an account of a sequence of historical events, from a theoretical account.² True, any causal explanation (in the ordinary sense of the term) involves theoretical claims that transcend the particular historical sequence of events. But theoretically oriented historical case analysis comes closer to a causal account than a cross-sectional variable oriented research; this for two major reasons: it enables us to establish the sequence of events and it helps to establish agency. Both of these greatly facilitate the identification of causal processes or, more precisely, of potentially causal processes that may or may not match hypotheses about causal conditions and sequences.

These considerations are not new except perhaps that they are here formulated in terms that also fit cross-national quantitative research. These

considerations explain why the radical neglect of historical research associated with functionalism and neo-positivism in the 1950s and 1960s appears now as a mere interlude. There was a long tradition of history-based social science, as illustrated by such names as de Tocqueville, Lorenz von Stein, Otto Hintze, and Max Weber, and this tradition has given rise during the past thirty years to a renaissance of comparative historical social science that in our judgment ranks among the major attainments of social analysis in the last generation.

ON HOLISM

Studying a given historical phenomenon in its context, studying it "as a whole," offers advantages for doing justice to historical particularity, but it may raise problems for systematic comparison. Goldthorpe's discussion turns this tension into a stark choice between a universalism in which country names are replaced by bundles of variables and a particularistic holism that makes any analysis across cases impossible. This alternative is misleading; it disregards the actual practice of comparative history. What we advocate—and practice—aims for an understanding of the case/country as a whole in order to facilitate detecting how social and historical factors combine in contingent ways to shape a given outcome.

The characteristics and factors to be examined are theoretically identified. Choosing which aspects and features to study is inevitably selective; there is no way of rendering a case as a whole in all its complexity without arbitrariness. On this we agree with Goldthorpe. We disagree with him—apparently at least—in our claim that due attention to the context of each case makes a different type of "operationalization" of concepts possible. Any operationalization involves hypotheses linking nominal concepts to indicators available for analysis. Cross-national statistical research settles on one standardized operationalization and takes inadequacies of fit, which vary across cases, into the bargain. Qualitative comparative historical research can give much closer attention to the match between evidence and theoretical conceptualization.³

These, then, are the major advantages we claim for historical analysis in comparative macro-sociology: it allows us to take account of historical persistencies and different constellations of major causal factors, it identifies sequences that are potentially causally relevant, it establishes agency, and it makes use of complex contextual knowledge in the operationalization of theoretical concepts. We will see that these advantages are of considerable importance for tackling the three problems under consideration.

CAPITALIST DEVELOPMENT AND DEMOCRACY

Our recent study on the impact of capitalist development on democracy (Rueschemeyer, Stephens, and Stephens 1992) will serve as a reference point

for several of our arguments in what follows; it thus requires a brief introduction.

At the outset we were confronted with a conflict between two bodies of research that differed in results as well as method. Cross-national statistical research, using numerical indicators for the level of socio-economic development as well as democracy, had demonstrated again and again—with different samples, different indicators, and different research designs—fairly strong correlations between development and democracy. By contrast, comparative historical research from Max Weber to Barrington Moore and Guillermo O'Donnell, using far more complex information about the historical trajectories of only a few countries, had come to much more skeptical conclusions about the prospects of democracy in cases of late development.

It was our judgment that cross-national quantitative research had established important empirical generalizations but that its causal interpretations were problematic. Comparative historical work, on the other hand, offered the elements of a more promising theoretical framework, a framework that had also proven useful in other comparative analyses of macro-phenomena—of class formation, of state structures, of welfare policies. We decided to tackle the explanation of the relationship between development and democracy that had been established by cross-national statistical research in a new way—by comparative historical analyses on as large a scale as we could manage. We included all advanced capitalist countries, the countries of South America, and a comparison of the Spanish-speaking countries of Central America with the English-speaking islands of the Caribbean. The historical analyses were guided by a theoretical framework that, based on our interpretation of past research and theory, identified the central questions, delineated the major concepts, and developed a core of hypotheses. Starting with the proposition that democracy is a matter of power, the framework focused on three clusters of power as the major causal constellations—on the balance of class power in society, on the structure of the state and of state-society relations, and on transnational structures of power.

The individual case analyses could, and did, engender additional explanatory hypotheses; and where these hypotheses contradicted propositions included in the initial framework, they could and did lead to modifications of the initial theory. "The result," we claimed, "is, on the one hand, a set of historical cases accounted for with a coherent theory and, on the other, a set of propositions about the conditions of democracy that have been progressively modified and are consistent with the facts of the cases examined as well as with the preceding research taken into account" (Rueschemeyer, Stephens, and Stephens 1992, p. 38; also Rueschemeyer 1991, p. 34). This is a precise formulation and a modest one. We will return below to Goldthorpe's claim that our attempt to explain the relationship between development and democracy failed because our procedure allowed too much leeway for the modification of theoretical propositions in response to the analysis of consecutive cases.

After his introductory comments on holism and historicism, Goldthorpe turns to the central thesis of his essay: He contends that, contrary to the claims of its advocates, comparative historical research is not superior to quantitative cross-national research in addressing the small *n* problem, the Galton problem, and the black box problem. In *Capitalist Development and Democracy*, we do indeed make this claim with regard to the small *n* and black box problems, and we would be willing to make the same claim with regard to Galton's problem though we do not address this question in the book.

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 might exceed the number of cases, making statistical 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 Stephens (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, as two of the collaborators in undertaking Ragin's Qualitative Comparative Analysis (QCA) identified over sixty characteristics which various theories had hypothesized 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.

In some cases, both qualitative comparative 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, it is assumed 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 the preferred independent variable (characteristic) in statistical analysis or *QCA*.

Other than this potentially misleading criterion, there are no other 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. Complementing the comparative method with historical analysis provides the researcher with such a tool. By uncovering agency and historical sequence, one can eliminate some potential causal variables and strengthen the case for others.

It must be admitted that our study, which compared a large number of cases on the basis of secondary historical accounts, is much better equipped to deal with the small *n* problem than most comparative historical studies. Indeed, it encompasses as many countries as many cross-national statistical studies, and the largest sample sizes are only twice that of ours. Comparative historical studies typically cover far fewer cases and, thus, though they gain by being able to consult with primary sources and by providing a more nuanced account of events in those cases, they sacrifice the advantage of having a larger number of cases which allows one to make fuller use of the comparative method to eliminate spurious factors. Our study combined the advantage of a substantial number of cases with the ability to study historical sequences and agency and to identify theoretical concepts in their varied contextual settings, but we had to sacrifice the use of primary sources as well as the nuanced attention to detail that characterizes the best comparative studies working on a few cases only.

Yet theory-oriented comparative historical work has yet another, more indirect way of dealing with the small *n* problem. The theoretical framework taps the results of earlier inquiries and thus indirectly eases the small *n* problem. "It is critical to fully appreciate this point because here lies one reason why the credibility of analytic induction is far greater than one could possibly justify with the few cases studied."⁵

GALTON'S PROBLEM

We agree with Goldthorpe that Galton's problem has been greatly exaggerated in the literature and that it can potentially be investigated by quantitative and comparative historical methods. However, we would contend that comparative history offers special advantages in dealing with Galton's problem as well. The reason is that historical research can trace a case over time and take full account of the way in which the characteristics that may or may not have been the result of diffusion are linked to their local context.

An example from *Capitalist Development and Democracy* will illustrate the point. In the book, we note that the correlation between democracy and British colonialism is a robust one. 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 (Chap. VI, also see pp. 280-281). 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 local 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.

The two most critical advantages of comparative historical research then, the ability to study the local context of critical characteristics and the knowledge of relevant historical sequences, are very efficient tools for settling the questions raised by Galton's problem.

THE BLACK BOX PROBLEM

There is little question that neither correlations between variables nor historical sequences as such can establish causation. In both modes of research, the solution is found in theory and in the close interplay between theory and the examination of the evidence. Comparative historical research derives, here too, its advantage from its focus on historical sequences and from the close interplay between theory and varied empirical evidence in the operationalization of the critical concepts. In examining historical sequences it furthermore reveals much about agency, which is virtually inaccessible in variable oriented quantitative research; and if agency does not tell the whole story of causation, it certainly is an important part of it.

Democratization in the West Indies, just discussed in regard to Galton's Problem, also demonstrates the superiority of comparative historical analysis in "opening up the black box," that is, in moving from statements about association to statements about causality. Knowing the historical record of the Caribbean islands, knowing what the British authorities did and how various

social forces pursued their interests in changing situations, we can confront these historical events with propositions about causal processes. In this way we come much closer to opening up the black box than cross-sectional statistical methods (or their qualitative counterparts) alone can ever do.

While, as Goldthorpe has pointed out elsewhere (1991), the historical "facts" underlying these narratives are based on incomplete data of varying quality, we are certainly far better off with knowledge of the historical record than simply with knowledge of a few data points or country characteristics. Moreover, most of the historical facts we rely on in our research comparison are ones which are disputed by few.

Goldthorpe correctly notes that in quantitative research, developing interpretations of the relationship would be aided by developing measures of the intervening variables, though he adds immediately that this would only mitigate, not eliminate the blackbox problem. However, it is not true that using quantitative analysis, we know what the "inputs" and "outputs" are, as Goldthorpe claims. We do not know this at all; their association could be spurious or the causal direction could be reversed and so on, as we pointed out above. For instance, in the case of quantitative research on democracy, it has been argued by various authors that democracy facilitates economic development. Consequently, without further information, we cannot even conclude that the central finding of this literature, that economic development is strongly related to democracy, can be unambiguously interpreted as demonstrating that economic development causes democratization. Interpreting quantitative measures of association requires theory and a close interaction between theory and the examination of evidence in the same way as these are required in comparative historical research. The question is which mode of research offers better access to the relevant evidence and which aids better the detailed interaction of theory and relevant evidence.

While refinements of the quantitative data may help to narrow the range of options, one must, we reassert, turn to comparative historical analysis "to open up the blackbox," to move to causal explanations. It is useful here to move from an abstract discussion to two concrete research problems, democratization and the development of the welfare state, to illustrate the potential and the limitation of refinements of data in contributing to an understanding of the phenomena in question.

In quantitative research on the development of democracy, the typical data set consists of a large number of countries (50 to 100), usually all of the countries in the world for which there are data or some large subset of it (e.g., all countries politically independent for a specified period of time, all developing countries). Generally, the data are a cross-section for some point in the post-war period.⁶ The most serious problem in these data is not the "small n—too many variables, too few cases" problem, though these issues do appear when the analyst attempts to examine subgroups separately (e.g., African countries only as in

one study; countries at the middle level of development which were democracies as of 1960 as in another). The key problems have rather been measurement of the independent variables due to the poor quality of the available data, questionable operationalization of variables, and exclusion of variables measuring important factors hypothesized in the literature to be related to democracy. Strength of the working class, strength of the middle class, state strength, income inequality, land distribution—these are just a few of the central concepts in the study of the social origins of democracy which are either inadequately operationalized in the cross-national statistical studies or are not measured at all. This is not due to sloppiness on the part of the scholars; indeed these studies often contain perceptive comments on the data limitations and ways to deal with them.⁷ The problem is that adequate data are not available for a sufficient number of cases; nor are they ever likely to be. For example, one measure of working class strength used in comparative welfare state studies is percentage of the labor force organized, which is available for all advanced capitalist democracies from 1950 to 1989 (Visser 1991). For third world countries, this figure is available only for a scattering of countries for any point in time, and the few figures available are of questionable quality. It is very unlikely that quality data on this key variable could be collected for the present and it is unimaginable that it could be collected for time points in the past.⁸

The problem of inadequate and incomplete data, then, takes another form in the case of cross-national statistical studies but it is at least as serious as, in fact it seems more serious than, in comparative historical research. The problem of inadequate data is in many cases linked to the inattention to historical sequences because of another measurement problem. The dependent variables and independent variables are typically measured at the same cross-sectional point in time. This is often very far removed in time from the critical events in the development of democracy in a large number of cases. For instance, the critical era in most European countries was the period 1870 to 1918, yet the inferences about causal processes based on cross-national studies are being made with data from 1960 or later.

We contend that our comparative historical study gets us much closer to the relevant historical evidence. Let us take the example of working class strength and democratization. In the case of the European countries in the 1870 to 1914 period, we have good data on union organization from the early 1900s and somewhat spotty data for some countries for several decades before that. We also have figures on votes for parties of the left, which we must, however, adjust for the extent of suffrage. From historical accounts, however incomplete and spotty, we know much about how centralized union movements were, how closely they worked with the labor and socialist parties, how unified the left was and so on; all of which bears on an overall assessment of working class strength. So we are in an excellent position to state that the working class movements were much stronger in 1910 than in 1870 and in

almost as strong a position to make comparative statements about in which countries these movements made the most rapid gains. If we then add information on the historical process of democratization, on what was the stance of the working class movement, who were their allies and enemies, and so forth, we are in a very good position to make statements about the effects of growing working class organization on democratization. True, these statements quite clearly involve broader theoretical assertions that transcend the cases at hand; but our evidence on historical sequences is quite precisely articulated with these theoretical assertions—contradicting them or, as was the case in this instance, confirming them.

Once we move to the Caribbean and Latin America, the “hard data” we can draw on are much spottier. However, we can make statements with which few historians disagree. For example, we do know that in both Central America and in the West Indies, plantation workers were not organized in 1930, that they rebelled in the 1930s, and that this, along with other events, led to the establishment of unions in the West Indies, but to the repression of the nascent working class movement in Central America. More precise measures would be helpful, but they are not essential to our interpretation because they are extremely unlikely to overturn our assessment of how strong the union movements were after the events of the 1930s had played themselves out. Similarly, every Latin American historian will agree that in the early post war period the Argentine working class was stronger than the Colombian which was stronger than the Salvadoran, but it would be hard to quantify just these three country assessments not to mention similar assessments for all Latin American countries we examine in the analysis. Nonetheless, our estimates of the relative strength of different movements are not arbitrary nor convenient ad hoc judgments; and they are transparently stated and open to criticism by other social scientists and historians.

Thus, while the main advantage of comparative historical analysis for opening up the black box lies in examining historical sequence, it also has the advantage that more theoretically important conceptual variables can be adequately “operationalized” and that they can be “measured” at the appropriate point in time. Given both of these advantages, it is hardly an exaggeration to say that we are in a *vastly* better position to make statements about causality from our comparative historical study than from any feasible cross-national quantitative study of democratization.

The data quality in the case of comparative welfare state research on advanced capitalist democracies is much, much better than in the case of democratization. We have reasonable and comparable measures of almost all central concepts in the literature and in most cases we have measures on an annual basis from the late 1950s on and for selected dates back to 1945.⁹ Moreover, this period, that is, the post World War II period, is the period in which the social policy provisions which account for most social spending were instituted. Thus, we

are measuring both the independent and dependent variables for the relevant time period. The primary problem in these data is the small n problem: the number of independent variables greatly exceeds the number of cases. Pooling cross-sections and time series data helps but does not solve the problem, because the increase in the number of cases is artificial. There is a sense in which there still are only eighteen rather than 600 cases. The fact that most of the variation explained is between the countries (about 1/2 using dummy variables for the countries) and not between the time points (about 1/5 using dummy variables for the years) makes this quite apparent. It is not surprising then that these data are plagued by multicollinearity among the independent variables to such a degree that it is not possible to enter some of them in the same equation much less sort out their relative (causal) effects. For example, union density, corporatism, and social democracy are so highly intercorrelated that it is impossible statistically to separate out their effects (and thus investigate whether union organization or corporatist bargaining arrangements are associated with social policy innovation independent of their association with social democratic governance and so on), much less move to some statement about causal relationships. Or, yet more perplexing, the aged proportion of the population is strongly correlated to social democracy ($r = \text{circa } 0.7$). Consequently, it is rare to find that they will both have strong coefficients when dependent variables are regressed on them and other determinants of welfare state development. For instance, Huber, Ragin, and Stephens (1996) have found that, net of ten other independent variables, aged population but not social democratic governance is strongly related to the public share of total health expenditure. By contrast, social democratic governance but not aged population is strongly related to a number of other indicators of the public funding of social services. The true believer in quantitative analysis would take these results as a demonstration of differences in the underlying causal processes. We are more skeptical. Since comparative historical research by Huber and Stephens uncovered relatively few instances of successful and important aged lobbying for health care reform (the United States being an important exception), we believe that this is an instance of multiple paths to the same outcome (mainly a Scandinavian-social democratic and a British) which is spuriously correlated to aged proportion of the population in the data. Thus, even if the data quality is extremely strong as it is in the case of crossnational data on advanced industrial democracies, making causal inferences from quantitative results is a hazardous process, and historical process knowledge can make a significant contribution in deciding between different theoretical accounts.

ON INDUCTIVE THEORY BUILDING AND TESTING

Goldthorpe offers a critique of inductive theory building and testing that is distinct from the "black box" problem and thus deserves separate attention.

The strategy we have called “analytic induction”¹⁰ begins with an explicitly formulated theoretical framework, in which questions, central concepts, and core hypotheses are clearly defined, and then proceeds to analyze a series of cases offering causal accounts of the outcomes in question. In these causal accounts of consecutive cases, new supplemental hypotheses may be developed in addition to the core theoretical propositions, and the core propositions—when contradicted—may have to be modified or rejected. The end result is a set of cases explained by a coherent theory. This theory has plausibility beyond the cases studied both because it was built initially on a critical analysis of past research and because it fits a tremendously complex and systematically analyzed body of empirical evidence from a number of historical trajectories.

Goldthorpe rejects this as a reasonable strategy. He maintains that the formulation of hypotheses must remain radically separate from their testing, that is, he insists on the classic neo-positivist distinction between the context of discovery and the context of validation. He claims that testing becomes in effect impossible if hypothesis development and hypothesis testing are intermingled. “If a theory is formed entirely inductively—without, so to speak, any deductive backbone—... it is hard to see how it can be genuinely tested at all... if a theory amounts to no more than an assemblage of inductions, the possibilities for ‘saving’ or ‘patching’ it in the face of contrary evidence are virtually limitless” (Goldthorpe, p. 15).

To begin with, neither our analysis in *Capitalist Development and Democracy* nor, say, Skocpol’s study of *States and Social Revolutions* is “entirely inductive” in character, a mere “assemblage of inductions” in Goldthorpe’s sense. The theoretical frameworks that open both investigations not only direct the subsequent case analyses but clearly rule out certain historical patterns, even if they do not constitute ready-made theoretical models that are then applied to all cases. Consequently, and quite contrary to Goldthorpe’s claim, these research strategies actually did not at all lend themselves to endless patching, and they ruled out other explanation attempts.

In *Capitalist Development and Democracy* we insisted at the outset that the prospects of democracy are contingent on three power clusters—relative class power, the structure of the state and state-society relations, and transnational power structures. For each of these we formulated specific hypotheses, which are not to be repeated here except for one of particular centrality: we claimed that the repeatedly established correlation between indicators of development and indicators of democracy was due to a shift in class power induced by capitalist development—a decline in the power of the landlord class and an increase in the organizational power of subordinate classes, in particular the urban working class. This we found to be supported in the vast majority of our cases. And our results starkly negate other explanations, among them (1) the thesis that it is the expansion of the middle classes that constitutes the primary link between capitalist development and democracy, or (2) the

modernization hypothesis that complex societies require a flexibility of government that only democracy can offer, or (3) the claim of both classic liberal and Marxist political theory that democracy is the creation of the bourgeoisie.

Even aside from this particular, if central hypothesis, our initial theoretical propositions were by no means endlessly adjusted. In fact, even the initial formulations fit a lot of cases. A lot of the changes just nuanced the original hypotheses—for example, concerning the role of landlords in the Caribbean and Australia—or they explained details of development not anticipated in the initial theoretical framework. Others did concern more genuine deviant cases—for example, about the working class in Argentina or the fate of democracy in Grenada. It is worthwhile noting, however, that in quantitative analysis, there is no attempt whatsoever to account for such cases. Given that we dealt with fifty cases we, too, could have left the few deviant cases without an interpretation that rendered them intelligible within the broader framework. By contrast, it is certainly true that in a comparative historical analysis of three cases, failing to explain one would be rather damaging for the study's persuasiveness; and if this explanation differs substantially from the original core theory, it would indeed have an improvised, ad hoc character.

Yet if Goldthorpe is wrong in assessing how the core theoretical propositions of *Capitalist Development and Democracy* fared in our comparative historical investigations, he is not wrong when he sees in comparative historical macro-sociology a methodological strategy that deviates from textbook prescriptions of method that are inspired by the neo-positivist program and primarily oriented to quantitative research. We did not lay out a fixed and complete system of hypotheses in advance of the investigation, but developed more fully detailed explanations in response to the individual cases, we tried to explain cases that did not fit our initial expectations, and we concluded with certain, if actually quite minor, revisions of the theoretical framework. The first question, then, is whether this is a legitimate procedure or whether it in effect constitutes a method of deluding ourselves and others. A subsidiary question is whether our "deviations" are just characteristic of comparative historical work or whether they are actually quite common in macro-social investigations, quantitative or qualitative.

Scholars seeking to explain important historical trajectories usually are not completely innocent of the relevant historical evidence and of earlier attempts at explanation. Thus, it becomes quite unrealistic to demand that hypotheses developed in qualitative case analyses be tested with fresh cases about which little or nothing is known in advance. At the same time, research that is guided by a systematic theoretical framework often generates quite substantial evidence that is not known in advance. If this framework does have a core of both clear-cut orientations for the development of specific hypotheses as well as a core theory, it is simply not possible to "patch things up" forever in the face of contrary evidence.

Even if a good deal of the explanations advanced pertain either to patterns known in advance or were developed as previously unknown historical evidence became available, the overall explanation may have much greater worth than the textbook injunction against "ad hoc" explanation suggests, especially if the different explanatory propositions fit into a coherent and more or less tight theoretical framework. Arthur Stinchcombe may be right when he suggests that it is easy to find at least three different theoretical explanations for any given correlation.¹¹ But this faith in a cornucopia of reasons is quite misplaced when it comes to explaining with a coherent set of theoretical propositions not just a single correlation but very complex patterns of historical sequences in different national contexts. In fact, we note that Goldthorpe has come to a quite similar conclusion: "While it is surely disappointing, it should not be thought a disgrace for sociologists to admit that they have not been able to develop a theory that will adequately account for their empirical findings" (p. 29, note 17).

As to the secondary question of whether "ad hoc" theorizing is found only in case-based qualitative research, we confine ourselves to a simple apodictic assertion which readers familiar with actual research will not find difficult to accept: It is rather naive to think that a similar back and forth of data examination and theory adjustment does not go on in quantitative work. Precisely because of the textbook injunction, however, research is simply written up as if that did not happen.

Before concluding, we should briefly comment on the issue of determinist vs. probabilistic theories. A determinist position appears to be forced on us if we ever take single cases as occasion for theory revision. But here, too, we contend, Goldthorpe formulates too stark a choice. A single case may rule out a theoretical proposition even if we do not assume that history is predetermined, that whatever happened had to happen. Whether a single case has such far-reaching implications depends both on the hypothesis in question and on the evidence of the particular cases. If we had assumed for example that peasants were inherently unable to organize collectively and thus were of little import for the push toward universal suffrage and were then confronted with the actual developments in Norway, Switzerland, and the northeastern United States we would be quite ill advised not to take the massive evidence from these three countries as a rejection of our hypothesis.¹² In other cases the outcome of a certain conflict may be at odds with a given hypothesis, but the evidence may suggest that it would be problematic to see this outcome as "necessary." Since in comparative historical research a "case" is more than a single dot on a scattergram, we are with this mode of analysis in a far better position to decide whether a deviant case must be considered conclusive evidence against a given hypothesis or whether it should be considered an instance of contingent developments that are more open for variation.

CONCLUSION

Max Weber took a rather skeptical and pragmatic view of methodological discussion: "It is possible to walk without knowing the anatomy of one's legs. Only if something is not in order does this knowledge become a consideration for walking."¹³ We do think that macro-sociology—of either kind—has some trouble walking, more so probably than Max Weber would have conceded; but we like the idea of giving priority to actual research over a methodological discussion based on stylized ideal assumptions.

We do maintain that comparative historical research has not only thorny problems, which it does, but also distinctive advantages. These advantages—above all its ability to consider evidence about historical sequence and agency and its capacity to identify the empirical counterparts of theoretical conceptualizations in a more complex and adequate manner—are of critical importance for all three problems under discussion.

In tackling the issues involved in the relationship between development and democracy, our comparative historical analysis struck a compromise, exchanging the opportunity of examining a very large number of cases for an almost complete reliance on secondary sources. In this, our strategy differed from other comparative historical work. Yet this choice enabled us to avoid the small *n* problem in its more severe form. And at the same time, the project could go substantially beyond what was possible with further quantitative research. Not only was additional data collection in the quantitative mode not likely to advance our knowledge substantially, more cross-sectional quantitative research could never have yielded the knowledge about sequence and agency that was won from comparative historical analysis and that allowed us to come closer to causal analysis.

We want to reiterate, however, that we recognize the strengths of good quantitative analysis. And we advocate a dialogue, or even a marrying, of the two. Despite the data drawbacks, the cross-national statistical studies on democracy were of great value to us. Many other comparative historical researchers ignored them much to the detriment of their own work.

NOTES

1. This has been our position all along: "Neither side has an obvious superiority in principle, and neither can be dismissed. Rather, each has made choices when confronted with a situation that did not allow obedience to all mandates of methodology—not even all the major mandates—at the same time. Each side had to pay for its peculiar strengths with equally characteristic weaknesses." Rueschemeyer (1991, p. 28); and Rueschemeyer, Stephens, and Stephens (1992, p. 32). Rueschemeyer (1991) is a revised version of an earlier draft of chapter 2 of Rueschemeyer, Stephens, and Stephens (1992). This chapter was written largely by Rueschemeyer but represents the view of all three co-authors.

2. In a footnote he accepts the idea that historical narrative itself can constitute a form of explanation. We want to make clear that this claim is not part of our position. It seems to involve a different meaning of "explanation."

3. True it has to deal with the danger of (unwittingly or intentionally) "adjusting" the operationalization to protect hypotheses from rejection; but then, all of these decisions are, in good research of this kind, laid open to scrutiny and if necessary dispute and rejection.

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. Rueschemeyer, Stephens, and Stephens (1992, pp. 37-38). Laitin (1995, p. 456) makes a similar point in his discussion of the recent King, Keohane, and Verba volume on qualitative methodology. Some studies incorporated in this indirect way may have used smaller and more numerous units of analysis, studying for instance the conditions of democratic decision making in unions and other voluntary associations. Though transferring results from the meso- to the macro-level may raise complex further questions, it is one possible way of dealing with the small n problem in macro-sociology. We did not take this route in our own research, but our theoretical framework included such references (see note 15, p. 312).

6. There are a few exceptions to this statement: Cutright and Wiley (1969) examined the relationship between development and democracy indices for 40 countries in four successive decades and established statistical associations between certain variables and constitutional change. An even closer examination of change is the event-history study by Hannan and Carroll (1981).

7. For example, see Bollen and Jackman (1985) on income distribution.

8. This can hardly be overstated. For example, in the course of writing a work on the Jamaican politics during the seventies, Stephens and Stephens (1986) attempted to collect comparative data on union organization in the West Indies, only to find that they could not even get accurate data on Jamaica where they were doing field research.

9. The dependent variable, almost always some measure of social expenditure as a percentage of GDP, was the most unsatisfactory of these measures. The most important recent innovation has been to measure social rights directly (Myles 1984; Esping-Andersen 1990) and at multiple points in time (Korpi 1989; Palme 1990; Kangas 1991). Another recent approach is to employ multiple measures of the dependent variable (Huber, Ragin, and Stephens 1993, 1996; Stephens, Huber, and Ray 1995).

10. We adopt a term coined for micro-sociology by Znaniecki two generations ago, but adapt the meaning to fit our methodological approach for current macro-sociology.

11. "A student who has difficulty thinking of at least three sensible explanations for any correlation that he is really interested in should probably choose another occupation" (Stinchcombe 1968, p. 13).

12. In fact, we considered from the beginning the hegemonic influence exercised by large landlords over peasants as the decisive factor for the limited role that peasants have played in most, but not all processes of democratization.

13. Weber (1913/1964, p. 139) as cited by Hennis (1994, p. 120).

REFERENCES

- Berg-Schlosser, D., and G. de Meur. 1994. "Conditions of Democracy in Inter-War Europe: A Boolean Test of Major Hypotheses." *Comparative Politics* 26 (3): 253-79.
- Bollen, K.A., and R. Jackman. 1985. "Political Democracy and the Size Distribution of Income." *American Sociological Review* 50.

- Cutright, P., and J.A. Wiley. 1969. "Modernization and Political Representation: 1927-1966." *Studies in Comparative International Development*
- Esping-Andersen, G. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.
- Goldthorpe, J.H. 1991. "The Uses of History in Sociology: Reflections on Some Recent Tendencies." *British Journal of Sociology* 42(2): 211-230.
- Hannan, M.T., and G.R. Carroll. 1981. "Dynamics of Formal Political Structure: An Event-History Analysis." *American Sociological Review* 46.
- Hennis, W. 1994. "The Meaning of 'Wertfreiheit': On the Background and Motives of Max Weber's 'Postulate'." *Sociological Theory* 12(2): 113-125.
- Huber, E., C. Ragin, and J.D. Stephens. 1991. "Quantitative Studies of Variations among Welfare States: Towards a Resolution of the Controversy." Conference on Comparative Studies of Welfare State Development sponsored by the Research Committee on Poverty, Social Welfare, and Social Policy of the International Sociological Association; Vuoranta, August 29-September 1.
- . 1993. "Social Democracy, Christian Democracy, Constitutional Structure and the Welfare State." *American Journal of Sociology* 99(3): 711-749.
- . 1996. "Political Power and Gender in the Making of the Social Democratic Service State." Delivered at the meetings of the American Political Science Association, New York, September.
- Kangas, O. 1991. *The Politics of Social Rights: Studies on the Dimensions of Sickness Insurance in OECD Countries*. Stockholm: Swedish Institute for Social Research.
- Katznelson, I., and A. Zolberg. 1986. *Working-Class Formation: Nineteenth Century Patterns in Western Europe and the United States*. Princeton, NJ: Princeton University Press.
- Lipset, S.M., and S. Rokkan. 1967. "Introduction." Pp. 1-64 in *Party Systems and Voter Alignments: Cross-National Perspectives*, edited by S.M. Lipset and Stein Rokkan. New York: Free Press.
- Palme, J. 1990. *Pension Rights in Welfare Capitalism: The Development of Old-Age Pensions in 18 OECD Countries 1930 to 1985*. Stockholm: Swedish Institute for Social Research.
- Myles, J. 1984. *Old Age in the Welfare State*. New York: Little Brown.
- Nettl, J.P. 1968. "The State as a Conceptual Variable." *World Politics* 20: 559-592.
- Rueschemeyer, D., E.H. Stephens, and J.D. Stephens. 1992. *Capitalist Development and Democracy*. Cambridge: Polity Press and Chicago: Chicago University Press.
- Stephens, J.D., E. Huber, and L. Ray. 1995. "The Welfare State in Hard Times." Paper delivered at the conference on the "Politics and Political Economy of Contemporary Capitalism," Humboldt University and the Science Center Berlin, May.
- Stinchcombe, A.L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace & World.
- Visser J. 1991. "Trends in Trade Union Membership." Pp. 97-134 in *OECD Employment Outlook*, July. Paris: OECD.
- Weber, M. 1964. (1913). "Gutachten zur Werturteilsdiskussion im Ausschuß des Vereins für Sozialpolitik." Pp. 102-139 in *Max Weber: Werk und Person*, edited by E. Baumgarten. Tübingen: J.C.B. Mohr.