7 Research Methods

Edwin Amenta
Alexander Hicks

Introduction

THE literature on welfare states and social policy has benefited from a wide-ranging debate across theoretical perspectives and methodological approaches. Scholars from different theoretical points of view have sought to examine claims across a vast stretch of empirical terrain. They have sought to develop and appraise hypotheses with methods ranging from the detailed examination of a policy sequence in one country to pooled cross-sectional and time-series analyses of social spending across all capitalist democracies. In this chapter we review some of these methodological approaches and address some of their achievements and drawbacks.

In this literature there has often been agreement over what is to be explained. The main focus has been on social spending, either overall or broken down into various types, such as transfers and services, or different functions, such as for health or old age. Social spending has been understood mainly in terms of its 'effort' or its amount as a percentage of economic activity (Wilensky 1975), with some attention to per capita spending (Pontusson 2005) and per-household income replacement rates (Allan and Scruggs 2004). There has also been attention to the adoption of major social programmes in the first half of the twentieth century (Collier and Messick 1975; Hicks 1999). Moreover, scholars in this area have sought to make sense of social policy 'regimes,' or overarching configurations of social policy (Esping-Andersen 1990; Castles 1997). In addition, scholars have sought to understand the retrenchment

of social policy since the 1980s, in terms of both alterations in programmes and reductions of spending (P. Pierson 1994; Hicks 1999; Swank 2002; Allan and Scruggs 2004). In the new century, scholarship has often broken past these bounds to examine more detailed questions and to examine programmes outside the usual definitions of social policy.

A wide variety of theories and hypotheses about social policy and welfare states have been developed and appraised. Theories have focused on modernization (Wilensky 1975), class struggle (Korpi 1983), political partisanship (Castles 1989), political institutions like states and party systems (Skocpol 1992; P. Pierson 1994), interest groups (Pampel and Williamson 1989), social movements (Amenta et al. 2005), cultural, world-societal influence (Strang and Chang 1993), and gender (Orloff 1993b; for reviews, see Amenta 2003; Hicks and Esping-Andersen 2005). In most instances, researchers have devised a similarly extensive range of empirical appraisals of theoretical claims. In scholarship on welfare states, there has been an unusual and fruitful dialogue between quantitative and qualitative historical research (Amenta 2003; P. Pierson 2007).

In our review of methods in the study of welfare states, we focus on what Amenta (2003) has termed 'causal research'. This sort of scholarship deploys self-conscious research methods to appraise theory and hypotheses in some significant way (Gerring 2007), or develops theoretical claims that are transportable in some fashion, such as in setting scope conditions on hypotheses (George and Bennett 2005), or both. The welfare state area as a whole has benefited from researchers addressing similar subject matter empirically from a variety of methodological approaches and often synthesizing approaches (Hicks 1999; Huber and Stephens 2001 a). The literature has been shaped by its focus on decisions made by states, the adoption of specific policies, or long-standing lines of state action (including individual programmes) that typically have consequences that are often easily measurable, such as the amount of spending devoted to individual programmes. It has led to historical inquiry regarding the adoptions and contractions of these policies and programmes and to quantitative and formal qualitative assessments of spending outputs. Given concern with outputs at the state level, the universe of plausible cases for examination has typically consisted of nations states. Given limited numbers of these, studies have mainly been of observational rather than experimental data employing convenience samples shaped by data availability.

In what follows we address these approaches. We do not provide an exhaustive review of all possible research but focus on examples, often from our own work. The latter spans in-depth historical analyses of a single country case to historical analyses of a few countries to Boolean QCA analyses across medium-N samples of countries and sub-national polities to cross-sectional and pooled cross-sectional and time-series analyses of countries and sub-national polities. We do this to show the variety of methodological work in the area and to highlight the advantages and disadvantages of different approaches. We conclude with suggestions for synthesizing, triangulating, and combining methods in order to minimize the disadvantages and maximize the advantages of different approaches.

Approaches to Causal Research

To situate different research approaches, we make broad distinctions between comparative and historical work. By *comparative* studies we refer to studies that address the experiences of two or more country cases (Rueschemeyer 2003), not one-country studies that make over-time comparisons or that simply situate empirical questions in a comparative context, and that make significant macro-level comparisons in the aid of causal inference. By *historical* studies we mean ones that include significant over-time variation in potential causes and place a premium on a deep knowledge of cases, path-dependent arguments, and a reliance on primary research (Amenta 2003). However, the main criterion for studies to fit our focus is that studies must take causality seriously in a double sense: to attempt to explain important welfare state developments by the appraisal of alternative hypotheses; to appraise, modify, or produce something at least partly theoretically transportable—a line of causal argumentation conceptualized so as to apply to cases or time-periods deemed analytically similar to those already studied (See Table 7.1).

Each of the categories includes both qualitative and quantitative studies. For instance, historical research may include primary document analyses of the development of policy in one country or quantitative time-series analyses of spending in a country or formal qualitative event structure analyses of policy adoption. Similarly, comparative and historical research may mean the qualitative analysis of the adoption of policy across a small-N sample of countries selected in a most-similar systems design or quantitative analyses of pooled time-series and cross-sections of social spending among capitalist democracies or event history analyses of programme-adoption across the world. Some methodological techniques may fall into more than

Table 7.1 Causal research according to methodological approaches

Historical Approach	Comparative Approach	
	No	Yes
No	Neither: Cross-sectional Quantitative Analyses of Within-Country Subunits; Medium- and Small-N Analyses across Subunits (1)	Comparative Only: Cross-National Small-N Analyses of Particular Periods, Formal Qualitative (QCA) or Cross-Sectional Quantitative Analyses across Countries (2)
Yes	Historical Only: Analytical Historical Case Studies; Time-Series Analyses of one Country (3)	Comparative and Historical: Small-N Comparative and Historical Analyses; Panel, and Pooled Time-Series Cross-Sectional Analysis (4)

one of the boxes; standard OLS cross-sectional regression can be used across countries, sub-national polities, or individuals, for instance. We discuss technique in terms of the methodological approach and the cell of Table 7.1 with which a technique is most closely associated.

Some studies may be comparative or historical, or both, but if they are not causal they fall outside the bounds of this review. These include the comparative studies that Skocpol and Somers (1980) refer to as 'contrast of contexts' and those that Charles Tilly (1984) refers to as 'individualizing comparisons', as well as historical case studies that employ social science concepts to interpret events, but fail to appraise alternative hypotheses or to develop theory seriously.

Neither Historical nor Comparative Causal Studies

It is possible, however, for a study to be neither historical nor cross-nationally comparative in the sense defined here, but still causal. The most notable examples are within-country analyses, using either large-N statistical techniques or formal qualitative ones analysing policy differences in federal polities across individual subunits (Amenta and Halfman 2000), small-N qualitative analyses across similar units, and studies of individuals (e.g., citizens, voters) analysed through survey data (Goodin et al. 1999). (See cell 1, Table 7.1.)

The main advantage in examining one country at one point in time is that many potential causes at the macro-political and macro-social level are held constant. In studies of sub-national polities, long-term conditions such as the overall political system and language are typically the same. So, too, are more short-term systemic conditions, such as the national political regime, state of the economy, and so on. An additional advantage is that the process by which data are collected is often similar across cases, as when national government agencies gather information about programmes administered at the sub-national level. These studies also often can yield tests of hypotheses that are sometimes difficult to examine cross-nationally, in that there may be more variation on some important causal conditions in subunits in one polity than across country-level polities. In the United States in the middle of the twentieth century, for instance, there was substantial variation among states in democratic political institutions, with some polities and areas of the country greatly restricting such basics as voting rights, and others extending these practices widely (Mayhew 1986; Amenta and Halfmann 2000).

However, the main advantages of such studies for adequate explanations of a given case also serve as a disadvantage in developing portable theory. The arguments may be closely bound to specific macro-social and macro-political conditions in ways that

are not easily understood and thus often make it difficult to sift out what is truly transportable in theoretical claims and what is specific to the case at hand. One critic (P. Pierson 2007) of quantitative political science research in American politics argues that it makes extensive general claims that are rarely formulated with explicit scope conditions or tested beyond the borders of the United States.

Comparative Studies

Strictly comparative studies (see cell 2, Table 7.1) encompass some small-N comparative studies and many early cross-national quantitative studies of social policy expenditures and policy adoptions. Small-N comparative studies also often are at least implicitly historical in approach; but some focus on delimited time-periods and in their argumentation deploy Mill's or similar non-dynamic methods of demonstration across comparative cross-sections. We refer to this as informal systematic comparison. In the comparative category we also place cross-national QCA studies (Ragin 1987; Hicks et al. 1995). Strictly speaking, however, QCA may be deployed on any sort of cross-section, including within-country polities (Amenta et al. 2005) and can also be used in ways that take time into account (Caren and Panofsky 2005).

Systematic Comparison

The most rudimentary of systematic comparative methods utilized in sociology is 'systematic comparison'. This method typically involves the identification of very strong and simple empirical patterns of traits across nations. The standard sorts of systematic comparisons tend to use Mill's methods of agreement and difference (Skocpol and Somers 1980), despite Mill's own warnings about the mechanical use of his methods on the observational data typical of the social sciences and central to research on welfare states. An instance of such a strong pattern can be found in Hicks (1999: 37), in the association between economic development and early welfare state 'consolidation'. With development measured as per capita real income over $2000 in 1913 (in 1980 dollars) and early welfare state consolidation meaning the adoption of at least three of the four major types of social insurance programmes—workers compensation, old age pensions, health care, and unemployment compensation—by 1920, Hicks finds that non-developed cases are always cases of no welfare consolidation. This pattern suggests that development is a necessary condition for welfare state consolidation. For association to be regarded as supportive of a proposition investigated by means of systematic comparisons, the pattern must be very strong and simple enough for detection by means of eyeballing. Here 'strong' means without, or almost without, exception, suggesting simple logical relations such as

'A is a sufficient condition for B', 'A is a necessary condition for B', or 'A is a sufficient and necessary condition for B'. The cases in which systematic comparisons can yield such clear results are rare, however. Moreover, theoretical claims usually are more complex, and only rarely will one-factor theories provide much analytical leverage.

Crisp- and Fuzzy-Set Qualitative Comparative Analysis

These circumstances promoted the development of qualitative comparative analysis (QCA), which extends and goes well beyond systematic comparison. Relying on set logic, QCA makes it possible to isolate conditions that are necessary and/or sufficient for specific outcomes (Ragin 1987, 2008). The goals and logics of using these analyses are somewhat different from those of social science research in which the explanation of variance is stressed. By using these methods, one selects a dependent variable and seeks to test or devise an explanation for it. Crisp-set QCA limits analysis to strictly dichotomous qualitative dependent and explanatory variables (for which elements are in a set with a probability of either 1.0 or 0.0). Instead of focusing on how much a given measure adds to explained variance, crisp- and fuzzy-set QCA both address conjunctural causation—the likelihood that two or more conditions must occur simultaneously to produce a result. They also address the possibility of multiple causation—that more than one conjunctural causal path will lead to a result. QCA can generate solutions that are accompanied by quantitative assessments of the strength of results, including statistical significance testing. QCA not only identifies conjunctions of causal factors; it focuses on the likelihood that given conjunctions of causal factors generate a given outcome.

Most extant work has been done through crisp-set analyses, and advancing beyond the example above can show QCA's value. Say we hypothesize that union strength, left-party strength, and Catholicism are all potent causes of welfare programme consolidation in 1920 and construct dichotomous measures of each of these variables (see Table 7.2); a visual assessment of the table suggests an imprecise pattern—yet applying the QCA algorithms produces this tidy solution:

Welfare Consolidation = [Early Union Strength AND Early Strong Left]OR [Early Union Strength AND Catholicism]

In this instance, early union strength is a necessary condition of both solutions. To produce welfare consolidation, however, also needed is

a strong left party or a Catholic country. If unions are strong in a Catholic country, a coalition will form to generate ample welfare programme legislation, even in the absence of a strong left party.

Fuzzy-set QCA (fsQCA) allows elements to have probabilities of inclusion in a set that vary from zero to one and offers several advantages over crisp-set QCA. FsQCA can address theoretical instances where cases do not completely fit a set and more mundanely can exploit the information in datasets that is lost when measures are reduced to dichotomies for crisp sets. In addition, fsQCA provides more estimates of

Table 7.2 Welfare programme consolidation in early democracies and proto-democracies (1920)

	No Welfare Consolidation; Non-Catholic	No Welfare Consolidation; Catholic	Welfare Consolidation; Non-Catholic	Welfare Consolidation; Catholic
Early Union Strength; Early Strong Left			Australia, Denmark, Sweden	Austria, Belgium, Germany
Early Union Strength; No Early Strong Left				Netherlands
No Early Union Strength: Strong Left	Canada, Norway, United States	France, Switzerland		
No Early Union Strength; No Strong Left				

the coverage of any given solution (the degree to which the solution or solutions overlap with the outcome set). It can better identify the consistency of the solution, meaning the degree to which cases with a given combination of causal conditions constitute a subset of the cases with the outcome. To date, however, the more precise fuzzy form of analysis has been seen only in research at the margins of welfare state research, such as analyses of the poor employment growth (Epstein et al. 2008) and the newspaper mentions of social movement organizations (Amenta et al. 2009).

The analytical advantages of QCA, crisp and fuzzy, are many. It allows scholars to address directly unusual circumstances that are often lost in less sophisticated analyses seeking to explain all variance, whether important and theoretically relevant or not, in a measure. Further, QCA encourages more complex theorizing and testing than quantitative techniques usually offer. It is far easier to think in terms of multiple conjunctural causes when there are means to analyse how well they fit relevant data. Finally, QCA seems advantageous in circumstances, common to welfare state research, where the number of (country) cases is too few for extended statistical analyses, but too many to permit conclusions based on inspection.

However, there are drawbacks to QCA. The many contending theoretical claims in welfare state research often lead to the simultaneous analyses of many measures. With QCA in any form, the number of potential explanatory measures (n) is highly limited; the number of combinations of these measures (2^n) quickly balloons beyond interpretive boundaries. An analysis of 10 causal measures will, for instance, yield 1024 combinations. (As we will show below, standard cross-national and time-series analyses suffer from more standard 'degrees of freedom' problems.) Also, analysing datasets with QCA

does not obviate the need of scholars to have a deep knowledge of cases. QCA in welfare state research typically must rely on observational datasets. Without a deep historical knowledge of the cases at hand, a researcher may miss the causal connections between concepts and measures vitiating findings. In crisp-set QCA, designing the 'truth table', a necessary preliminary for analysis, is a task with steep knowledge requirements (Ragin 1987). Decisions about break points and degree of membership in categories of fuzzy sets also depend on such substantive knowledge (Ragin 2008). (For the most complete discussions and software, see Ragin 2008; Longest and Vaisey 2008.)

Cross-Country Statistical Analyses

Where the number of available country cases is large enough for the operation of statistical inference, social scientists often employ it, even though cases and their variable traits are not typically randomly distributed. However, the analyses of data lacking explicit temporal dimensions can still be revealing. Operationally, causal claim can be bolstered if associations are not statistically spurious, changes in explanatory variables precede changes in the outcome variable and the explanation relies on an overarching theoretical argument or a specification of plausible mechanisms translating changes in explanatory variables into changes in outcome ones. The typical tool used in comparative statistical analyses of welfare states is OLS multivariate regression (Cutright 1965; Cameron 1978; Myles 1984; Hicks and

Swank 1984; Pampel and Williamson 1989; see also the panel analyses of Hicks and Kenworthy 1998, 2003).

These sorts of regression analyses were influential and important in adjudicating debates and advancing knowledge in the early years of research on welfare states, establishing some basic points about welfare state development in the expansionary immediate post-World War II period. Researchers were able to show fairly conclusively that the level of economic development was crucial to the explanation of welfare state effort at all income levels (Cutright 1965; Wilensky 1975) and among poorer countries (Collier and Messick 1975). However, economic development level did little to explain variations in welfare state development among the rich capitalist democracies (Myles 1984; Hicks and Swank 1984), where political factors apparently offered better explanatory leverage (Castles 1982 *b*; Myles 1984).

An advantage of this sort of research, as with all quantitative work, is the ability to examine concurrently large ranges of theoretically indicated hypotheses against alternatives. However, simple cross-sectional analysis has largely been abandoned by scholars, because of the relative inability to appraise large numbers of measures due to degrees of freedom issues across selective and theoretically important groups like rich capitalist democracies. Further, simple cross-sectional regression cannot exploit case-based knowledge or yield conjunctural findings as well as can formal qualitative methods like QCA. For the most part, scholars other than those in the first generation of researchers have concentrated their attention on the numerous data points and historical specificity provided by pooled cross-section and time-series analyses (treated below).

end p.112

Historical Studies

Several types of single-country case studies are considered 'historical' in our analysis of causal research (see cell 3 of Table 7.1). We address, first, historical narrative work that gains empirical leverage by examining causal propositions by considering historical developments and sequences. The second main type includes quantitative time-series analyses of datasets within one country. A third type includes event structure analyses, which typically examine sequences of events within one unit, often a country (Griffin 1993; Isaac et al. 1994).

Historical Studies Proper

Historical causal studies deploy a scholar's deep historical knowledge, documentary research, and the examination of specific episodes of change to appraise current hypotheses or develop new ones. Historical knowledge makes it possible to identify key instances of relevant variables and events, and allows the detailed examination of causal mechanisms. These analyses typically focus on reasons behind key events such as policymaker goals behind major changes in policy (Skocpol 1992; Castles 1989). Often scholars can identify through documentary evidence what different key actors were proposing and what was likely to occur in terms of policy had some historical event or intervention not taken place. Often these empirical demonstrations of propositions are used to generate portable explanations, historical instances of what Strauss (1987) called 'grounded research' (see also Eckstein 1975).

Scholars have claimed that there are several standard disadvantages attached to this sort of research. In the standard view, because they provide an N of one, historical studies are seen as limited to the development of hypotheses or a means to reject specific general hypotheses (Rueschemeyer 2003; King et al. 1994). Also, as mentioned in the discussion of within-country analyses, many of the aspects of the context seemingly 'controlled for' in these studies may be relevant contextual causal factors, yet taken for granted and left un-theorized.

However, scholars have recently noted many advantages of the historical approach. One of these is its ability to examine theoretical mechanisms and the details of causal arguments. This sort of analysis is frequently referred to as 'process tracing' (George and Bennett 2005; Gerring 2007). If, for instance, a partisanship theory expects key programmes to be adopted under specific types of regimes or to be championed by specific sorts of actors, a historical scholar can examine just how much these conditions mattered in a given episode of policy making. Similarly, if a theory anticipates legislative or other key action to take a specific sequence, a scholar with deep knowledge of the case can probe these theoretical mechanisms with subtlety. In addition, the disadvantages surrounding contextual similarity are mitigated to some extent by the fact that in historical studies relatively long time-periods may exhibit variation of contextual conditions. Moreover, most such studies make

end p.113

comparisons across policy areas and programmes, as well as across periods of activity, inactivity, and retrenchment, and these sources of variation make possible further hypothesis testing (Amenta 1998). Most of these methodological moves expand the number of observations beyond 'N equals 1' (King et al. 1994).

There are advantages that go beyond simply exploiting the various types of historical information in a country case. In addition, historical

one-country research can generate and appraise more sophisticated theoretical claims that take into account time order sequences (P. Pierson 2000) and configurational and multiple causes (Ragin 2008). Social scientists working on a case through primary materials can avoid the biases of reliance on secondary research, as in most qualitative comparative historical analyses (Lustick 1996). Historical scholars may also contribute to quantitative research by identifying new sources of data and can often build more valid indicators than can scholars mining data from standard sources. Most of all, historical studies are often asking the kinds of big questions that are simultaneously theoretical and historical (P. Pierson and Skocpol 2002), such as why the United States did not develop a welfare state on the European model when some theories would expect that to have happened. These questions are of both public and social science interest and cannot be addressed simply by way of expanding observations. A deep causal understanding of one case, analysed social scientifically, may provide the best building block for further theoretical argumentation (Mahoney 2000).

Event Structure Analysis

Another means to analyse data over time is event structure analysis (ESA), a formal qualitative tool that forces scholars to be explicit about their counterfactual reasoning. ESAs are designed to provide narrative causal accounts of particular sequences of events (Griffin 1993), culminating in an event of importance. These may include those leading to the enactment of social programmes (Isaac et al. 1994). In providing specific interpretations and causal accounts of key events, these analyses provide aid in the development of theory that is historical in a specific social science sense. That is, these analyses are based on the historical institutionalist insight that *when* something happens in a sequence may be causally important as to why something happens (P. Pierson and Skocpol 2002; Mahoney 2000). Specifically, the standard event structure analysis uses the ETHNO programme developed by Heise (1989). This programme induces a researcher to answer a series of questions regarding a specific causal account, forcing consistent reasoning and allowing for the possibility of replication. It is a kind of process tracing that has been specified to the inductive development of causal accounts. Although the models developed from such analyses are most valid as a causal interpretation of a case, the most useful result is one in which a series of potentially portable middle-range and time-ordered theoretical claims are developed.

Although ESA is a promising tool, its use in social policy analysis is still only emerging, and it must overcome a few hurdles. For the most part, event structure analyses require valid and reasonably complete factual accounts of key events or the

means to reconstruct them. Only a deep knowledge of the case materials can aid in understanding where timing might and might not matter and in addressing the many counterfactual questions inherent in this type of formal qualitative historical research. Although these analyses are replicable, it is likely that different scholars will answer the questions about causality differently even for the same case rendering rigour and persuasiveness especially important. But used appropriately ESA can aid in both the causal interpretation of specific cases and the development of more comparative causal hypotheses that address time order explicitly.

Time-Series Analysis

Another key type of historical research concerns quantitative analyses over time within one country, or the statistical analysis of times-series data (Janoski 1992). Within- country across-unit designs that pool cross-sections and time-series data are non-comparative, but they remain historical in the sense used here (Amenta et al. 2005). Although these studies are not necessarily historical in the sense that the author has a deep knowledge of the cases at hand, they often make important claims about the importance of the working of over-time processes within structurally determined time-periods (Isaac and Griffin 1989).

Time-series studies are characteristically studies that employ multivariate statistical techniques, and that thus have the standard advantage of being able to examine several hypotheses simultaneously. Estimation employing the core statistical procedure, termed 'generalized differences', can entail problems in particular circumstances, but solutions for these are available (see Ostrom 1978 on inefficiency; see Gujarati 2003 on lagged dependent variables and inconsistency; Greene 2000). These analyses, however, also require some historical knowledge: mainly in delimiting the time-periods for analysis, and identifying the beginning and end of a homogeneous process such as an era of expansion or retrenchment.

However, time-series analyses have not only some of the standard analytical disadvantages of one-country research, they also miss the benefits of strictly historical work. The questions addressed in time-series analyses are usually at one remove from the policy-making decisions at the centre of historical work and usually address important but limited aggregates such as spending. These analyses also typically suffer from a small-N problem, as identifying coherent time-delimited processes also limits the number of cases for statistical manipulation, though sometimes sub-national polities can be deployed to augment the number of observations (Amenta et al. 2005). Also, however, these analyses have difficulty in addressing time-invariant explanations and factors, such as political institutions and structure of labour movements. For these reasons, scholars often seek to bring across-country evidence to bear, such as devising methods to

compare and analyse time-series parameters across countries (Western 1998), using these parameters as descriptive evidence to be explained. More generally, scholars have sought to harness the potential advantages of this research to designs that pool time-series across countries, which we address below.

Comparative and Historical Research

As with the previous modes of research, comparative and historical (cell 4, Table 7.1), research includes both qualitative and quantitative studies and approaches. Notably, we include here historical analyses of several country cases and the statistical analysis of a cross-section of country cases deepened by the examination of time-series data for each. On the qualitative side, we have what might be termed 'classical comparative and historical research'. On the quantitative side of comparative and historical research there are, most notably, pooled data analyses of cross-sections and time-series. These are typically also carried out across capitalist democracies, usually, however, seeking complete coverage of cross-sections, and focused on the post-World War II period.

Classical Comparative and Historical Research

Classical comparative and historical research at its most systematic is often conducted like comparative research, for example employing systematic comparison by means of the Millsian method of agreement and difference. It typically employs 'most similar systems' designs (Przeworski and Teune 1970), in which characteristics are 'controlled for', or become part of the scope condition of the claims, such as advanced capitalist democracies, or liberal welfare state regimes (see also Gerring 2007). Many key qualitative works make small-N comparisons across long stretches of time in order to appraise and develop hypotheses about some aspect of social policy (Heclo 1974; Baldwin 1990; Steinmo 1993; Immergut 1992a; Orloff 1993 *a*; P. Pierson 1994; Amenta 1998). Classical comparative and historical researchers have been able to address a wide range of theoretically and historically important questions: why did social policy take off when it did and why did it become so prevalent? Why did some countries lead and why did some others fall behind in different phases of the development of social policy? Why did some states adopt distinctive forms of social policy? By situating the experiences of different countries against the group portraits, these researches brought to light historical anomalies and puzzles to solve.

Comparative historical research has provided most of what we know regarding the early adoption of social policies among the more economically developed countries. It has also helped to address debates between political institutional approaches based in Weberian and Tocquevillian theory and focusing on the structures of political institutions and political organizational approaches based in Marxian theory and focusing on the political organization of social groups, notably the labour movement (see review in Amenta 2003). From a focus on left-wing or social democratic party rule, scholars have moved on to consideration of the role of the right-wing parties (Castles 1985), farmer-labour political coalitions (Esping-Andersen 1990), expert-labour alliances (Orloff 1993 *a*), and Christian democratic rule (Huber and Stephens 2001 *a*). Comparative and historical scholars have also gone on to build more

theoretically synthetic or configurational arguments, combining the structural strengths of institutional claims with the strengths of claims based on political identities and action (Skocpol 1992; Amenta 1998; Hicks 1999; Huber and Stephens 2001 *a*).

Classical comparative and historical research has most of the advantages of historical research on one country, mentioned above. It also has the additional advantage of being able to compare similar trajectories of countries and to pinpoint and explain divergences in policy development (Rueschemeyer 2003). This method has some of the well-known disadvantages in appraising hypotheses also discussed in the treatment of historical qualitative work. And despite the addition of cases, classical comparative and historical research, with its steep informational requirements, rarely addresses complete populations of theoretically relevant cases. Thus researchers must rely on secondary research and its attendant biases (Lustick 1996). Quite possibly this research works best in most similar systems designs such as across 'social democratic' welfare states or 'English-speaking' countries where it is possible for researchers to engage in small-N designs without having to rely as greatly on secondary research.

Classic comparative and historical scholars have also been free to rethink what social policy meant and to deepen the concept. Comparative and historical scholars have been able as well to appraise theoretical arguments by addressing social programmes other than the ones prominent in quantitative work. Among the possibilities were veterans' benefits (Skocpol 1992), education (Heidenheimer 1981), taxation policy (Steinmo 1993), housing policy (P. Pierson 1994; Castles 1998 *b*; Bonastia 2000), economic policy (Hall 1986; Weir 1992), and work programmes (Amenta 1998). In this process, comparative and historical scholars have devised new questions and have opened up new research agendas, helping to develop and refine theoretical argumentation (see review in Amenta 2003). Another way to develop the research agenda and advance theory has been to split the concept by taking the developmental *phase* of social policy seriously, entertaining the possibility that different phases of social policy have different determinants, as in the case of the retrenchment of social

policy (P. Pierson 1994), which is a more difficult process than adopting one and depends crucially on processes set in motion by the nature of the policy in question. This argument has been deemed appropriate to explain social policy developments since the 1980s (see also Huber and Stephens 2001 *a*; Swank 2002). By this time most systems of social spending had been completed and expanded—had become 'institutionalized'—and bids to cut them back were taken up in force by many political regimes.

In addition to the conceptual 'splitting'—with its attention to possibilities of causal heterogeneity across blocks of time or sets of cross-sections—classical comparative and historical research has been at the centre of broader conceptualizations of social policy or ideal types that characterize policies as wholes. Building on previous models of social policy, Esping-Andersen's (1990) 'welfare state regimes' address social policy's influence on labour-market relations. New conceptualizations of social policy have also been provided by feminist scholars (see Chapter 17), many of whom work in a comparative and historical mode (Skocpol 1992; O'Connor et al.

1999). The main responses have been to modify Esping-Andersen's types or to replace them with gender-based policy regimes.

In short, classical comparative and historical research has many of the benefits of historical studies proper, with the added advantage of throwing a case into relief vis-à-vis one or more largely similar cases. This makes it possible to ask new and specific questions in analysing diverging trajectories in social policy. These analyses can be used to appraise theories and arguments and to ascertain how far they extend to different processes and outcomes outside a particular domain of welfare state research, as well as to create or enrich theory for appraisal across other cases. However, typically the arguments cannot be appraised fully with the cases at hand. Knowledge requirements are steep for these analyses and adding cases without the requisite background can lead to a thinning of analyses, issues of validity, and the loss of the advantages associated with historical research proper. Nonetheless, these analyses have played a central role in driving the agenda of research on welfare states (see Amenta 2003).

Statistical Analyses of Pooled Cross-Sections and Time-Series

The analysis of pooled cross-sectional and times-series data has been the method of choice of most sophisticated quantitative analyses of social policy over the last 20 years (see Hicks 1999; Huber and Stephens 2001 *a*; Iversen and Cusack 2000). Pooled data involves measures arrayed across both time and space, and analysing these data helps to overcome some of the shortcomings of cross-sectional and time-series analyses separately. Pooled analyses can address variables acting over time, such as changes in partisan regimes, as well as temporally inert, structural variables, like policy regimes or political structural arrangements.

Pooled data analyses also solve small-N problems entailed by analyses of time-series of limited per-nation length and cross-sectional domains of limited numbers of nations (e.g. the approximately twenty advanced capitalist democracies). The solution is pooling time-series and cross-sections. In addition, pooling helps to illuminate stable differences among countries—as did early cross-sectional research—in addition to exploring the dynamic processes and changes in social spending efforts, the emphasis of time-series analyses. Key to this enterprise has been the excellent and voluminous data collected in the post-war period on these countries by various international organizations, especially the Organization for Economic Co-operation and Development (OECD) and the International Labour Organization (ILO). These data have been augmented by individual scholars (e.g. Huber and Stephens 2001 *a*; Swank 2002). Although there are several problems with the estimation of these models, including over-time and spatial auto-regression, heterogeneous regression intercepts and slopes over time and/or space, and heterogeneous error variances across time and/or place, many means of addressing these estimation challenges are available (see Hicks 1993; Beck 2007; Plümper and Tröger 2004; Hicks and Freeman 2009).

Although data and analysis of welfare states have been largely confined to twenty or so long-standing, OECD democracies, impressive efforts have been made to extend data to Latin America (Huber et al. 2006; Brooks 2009) and beyond to East Asia and Eastern Europe (Haggard and Kaufman 2008). Moreover, studies have begun to address the complexities of transnational as well as sub-national federal contexts (Amenta and Halfman 2000; Obinger et al. 2005 *b*).

In short, pooled regression analyses across rich democracies in post-World War II period have helped to answer many questions and to resolve debates about spending efforts in the periods of welfare state growth and retrenchment. Moreover, resolution of kindred debates is extending beyond the rich long-standing democracies to new parts of the world and transnational contexts.

To analyse statistically the adoption of programmes, as well as legislation regarding retrenchment, similar data have been used to predict nominal or qualitative outcomes, involving techniques falling under the rubric of 'event history analysis' (Usui 1994; Hicks 1999; Hicks and Zorn 2005; for overviews and technical treatments, see Allison 1974; Box-Steffensmeier and Jones 2004; Kleinbaum and Klein 2005 *a*, 2005 *b*). These analyses are often undertaken across all countries, notably addressing the issue of the adoption of policies outside the

domain of rich capitalist democracies, filling gaps in knowledge. However, given their wide reach, these analyses are often beset with even greater problems of missing data in respect of potentially relevant measures.

Conclusion

Research on welfare states and social policy has addressed all manner of methods. Variability in the availability of forms of data has led to a kind of division of labour. Hard and systematic data suitable for quantitative studies existed only for the post-World War II periods of expansion and retrenchment, while less complete information of this sort was available for the periods of adoption, consolidation, and completion (Hicks 1999). Quantitative comparative researchers mainly analysed data from the 1960s and beyond, and comparative and historical researchers took charge of the first half of the century. There has been a great deal of interesting work as well, however, on the 'off-diagonal' cases, and these studies were often sites of innovation and spurs to analysis of the other variety (see Amenta 2003). Paul Pierson (1994), for instance, opened a new line of thinking and research on retrenchment, providing hypotheses later addressed by quantitative researchers. The quantitative paper by Collier and Messick (1975) cast doubt on the modernization thesis with respect to the adoption of social policy and spurred comparative and historical work, and work by Hicks et al. (1995) applied QCA analyses to policy adoption in the first half of the twentieth century.

The development of the area was accelerated by the open-minded methodological outlook of many prominent researchers. Few quantitative researchers derided the work of comparative and historical researchers as lacking in rigour. Few comparative and historical researchers saw the work of the quantitative scholars as simplistic and lacking in depth and validity. The tone was set early on, with Gaston Rimlinger (1971) employing the gold standard of quantitative studies, social spending 'effort', to situate his path-breaking comparative and historical investigations. The willingness and ability of researchers to work in different modes was key. Francis Castles, Gøsta Esping-Andersen, Alexander Hicks, Evelyne Huber, Torben Iversen, John Stephens, Theda Skocpol, and Duane Swank comprise some of the scholars proficient in one type of methodology, but willing or driven sometimes to employ others.

Given the various strengths and weaknesses of the different methodological approaches, as indicated above, it is useful for scholars to employ more than one. Scholars of social policy have combined methods and triangulated them, reaping advantages, while minimizing the disadvantages of any single approach. Studies have been able to address large questions about differences in timing or trajectories or outcomes in social policy, by examining a few cases or one case in a comparative context, by developing and tentatively appraising relatively complex arguments and by sometimes employing multiple and conjunctural causation, sometimes involving mechanisms of process. Standard quantitative techniques have difficulty in assessing these more complex questions, much as standard comparative and historical work can provide only rudimentary tests of more general hypotheses.

There have been many examples of syntheses and triangulation of methods in the welfare state literature. Evelyne Huber and John Stephens (2001 *a*) address the rise of welfare states and efforts at retrenchment over the last decades with pooled time-series and cross-sectional regression analyses and detailed case histories of different types of highly developed welfare states in examining the development of social policy over long periods. The latter technique is used in order to get around the short-term biases in regression analyses and to closely examine critical periods of policy change. Other examples include work by scholars who employ different techniques on the same subject matter across different works (Skocpol 1992; Skocpol et al. 1993; Pampel and Williamson 1989; Amenta 1998; Amenta and Halfmann 2000).

These scholars have understood that different approaches had advantages and disadvantages, and exploited the advantages of each to allow greater progress than could be achieved by one or another approach. This outlook has helped advance the field tremendously. Preserving this outlook should bring still greater advances in the future, as both qualitative and quantitative approaches increase in sophistication.