Introduction to the 2014 Edition

ORIGINS OF THE APPROACH

More than a quarter of a century has passed since *The Comparative Method* was first published. The occasion of its republication provides the opportunity to reflect on the work and its impact. *The Comparative Method* formalizes the logic of case-oriented comparative research and elaborates an analytic approach centered on its principles. This new analytic approach, called Qualitative Comparative Analysis (QCA), comprises a set of strategies and techniques that both bridge and transcend the qualitative-quantitative divide in social research.

The development of *The Comparative Method* and QCA followed from several methodological issues that I confronted in the 1970s and 1980s. I was trained as a quantitative social scientist but became increasingly frustrated with the limitations of this approach. Early in my graduate studies I read Barrington Moore's *Social Origins of Dictatorship and Democracy* (1966). This book proved to be highly influential in the development of my approach because it offers an excellent in-depth comparative analysis of a limited number of cases, focused on the combinations of antecedent conditions linked to specific, large-scale historical transformations. A key feature of this work is that its argumentation resists the language of independent variables and their net effects. Here was a highly influential work of substantial social scientific merit that could not be squared with the dominant quantitative discourse.

The methodological challenge I faced was to formalize an approach that would enable researchers to systematically integrate within-case and cross-case analysis, as Moore had accomplished. An important concern was remaining true to the nature of qualitative argumentation, with its key focus on the question of how things happen. Addressing the "how question" should be the starting point of any technique that seeks to use comparative case analysis. In *The Social Origins of Dictatorship and Democracy,* Moore answers the "how question" by focusing on the different combinations of antecedent conditions linked to divergent political outcomes.

A second factor that influenced my effort to combine qualitative and quan-

titative approaches was a collaborative research project on the Romanian peasant revolt of 1907 with one of my advisers, Daniel Chirot (Chirot and Ragin 1975). Our concern in this project was to investigate the combination of conditions linked to the outbreak of peasant revolts, using quantitative evidence. Building on the macrosociological tradition in research on revolutions, the paper identified two models that it tested using historical data on the Romanian peasant rebellion. The pivotal model built on work by Eric Hobsbawm (1959), Charles Tilly (1967), and Moore (1966) and stressed the tension between the survival of peasant traditionalism and the intrusion of capitalist market forces. In this paper I pursued John Stuart Mill's notion of chemical causation (Mill [1843] 1967), arguing that causal conditions often must combine in order to generate qualitative change, and I operationalized this argument via the statistical analysis of interaction effects. The paper shows that the model does explain the Romanian revolt and that the important condition explaining spatial variation in the intensity of the rebellion was the interactive effect of peasant traditionalism and the penetration of market forces in agriculture.

Identifying interaction effects as a key explanatory strategy started a five-year journey of looking systematically into interaction models. For a variety of reasons, this increased attention to interaction effects within the context of statistical modeling also produced increased frustration. The results of a test for interaction are shaped in part simply by the metrics of the component variables and their degree of correlation. Substantial colinearity between an interaction term and its components is a recurring issue. When higher-order interaction terms are considered, colinearity is exacerbated considerably, and researchers often have their choice of which interaction effects to report. I concluded that working with interaction effects, especially three-way and higher-order terms, was an extremely fragile enterprise. However, this was one of the areas I wanted to explore further, in order to assess the conjecture that there could be mixtures of four, five, or six conditions generating a qualitative change. This way of viewing social phenomena was much more akin to the historical analysis of macrosocial phenomena. I worked for years trying to make interaction models, especially models with higher-order interaction effects, more robust. In the end, I concluded that this was not a fruitful path and that alternate techniques had to be developed.

This concern and related problems prompted a search for a new approach. My aim was, in essence, to formalize some of the practices that are common (and often implicit) in case-oriented research. In order to formalize how dif-

ferent conditions combine to generate a qualitative change, I turned to books that I had read as an adolescent on Boolean algebra, set theory, and switching circuits. At about the same time, I moved from Indiana University to Northwestern University, which was then more open to qualitatively oriented research. In this environment it was possible for me to experiment with new analytic approaches, and I presented my first lectures on my emerging methodological ideas in seminars at Northwestern.

This exploration of alternate analytic foundations culminated in the development of QCA, and in 1987 *The Comparative Method* was published. It was intended not as an end product of a process of methodological innovation but more as a first step in a work in progress that was ready to be communicated to a wider audience. The approach set forth in this work has four main components.

First, it emphasizes the case-based nature of comparative research, an emphasis that demands that each case be treated as a complex entity (a whole) and that its integrity as a case be maintained in the course of the analysis. Different parts of each case should be understood in relation to one another and in terms of the whole that they form together. The organizing idea is that the parts of a case constitute a coherent whole and that the effects of variables should be assessed in the context of the case and not detached from it. In order to operationalize this idea, cases are represented as configurations of conditions. The essence of the analytic approach is to link configurations of causally relevant conditions to outcomes.

Second, the approach is comparative in the sense that it enables researchers to explore similarities and differences across comparable cases by pooling similar cases and comparing them as configurations. The analytic device that allows this is the truth table, which displays the data in a matrix of logically possible configurations of causal conditions. By bringing cases together in a truth table, researchers can assess which ones display identical configurations of causal conditions and which differ on one or more causally relevant conditions.

Third, explanatory models are developed in an iterative manner, so as to facilitate a dialogue between theory and evidence. A key mechanism for developing an explanatory model is the presence of contradictions. Contradictions occur when an identical configuration of conditions is linked to both the presence and the absence of an outcome. Contradictions are revealed through the transformation of a data matrix into a truth table. If a truth table reveals contradictions, they should be resolved, primarily by identifying omitted

causal conditions. Hence, the development of an explanatory model goes hand in hand with resolving contradictions. This back-and-forth process of including and excluding theoretically and empirically relevant conditions in a model until it has been identified with no or only a few contradictions is the key mechanism for developing an explanatory model.

Fourth, as a result, QCA allows for the assessment of multiple conjunctural causation, which implies that (1) most often, it is a combination of conditions that generate an outcome; (2) several different combinations of conditions may produce the same outcome; and (3) a given condition may have a different impact on the outcome depending on context. Hence, QCA implements a context-specific notion of causality. This allowance for greater causal complexity also implies that a causal condition may have opposite effects, depending on context. As a result, by using QCA, the researcher is urged not to "specify a single causal model that fits the data best" (the standard practice using statistical techniques) but instead to "determine the number and character of the different causal models that exist among comparable cases" (Ragin 1987 : 167).

The version of QCA that this book presents was developed for use with dichotomous (crisp-set) conditions and outcomes. Soon after the publication of *The Comparative Method,* I started work on the development of fuzzy-set analysis as a complement to crisp-set analysis. This line of inquiry led to the publication in 2000 of *Fuzzy-Set Social Science,* followed by the publication in 2008 of *Redesigning Social Inquiry: Fuzzy Sets and Beyond.* In fact, the development of the fuzzy-set approach and new software for the analysis of fuzzy-set relations (fsQCA) predated many of the debates focused on the limitations associated with the use of crisp sets.[1] The addition of fuzzy-set analysis to QCA set the stage for the extension of the approach beyond the domain of case-oriented macrocomparative inquiry.

A NEW APPROACH TO SOCIAL RESEARCH

My initial goal in writing *The Comparative Method* was to formalize the logic of comparative analysis, as practiced by case-oriented comparative researchers, and to unveil the set-theoretic foundations of comparative analysis. The key step was making explicit the configurational thinking that undergirds comparative inquiry. This also provided a way to address the many communication breakdowns that occur when quantitative and qualitative comparativists venture out of their narrow methodological confines and

TABLE A: Conventional versus Alternate Template for Social Research

Conventional Template	Alternate Template
Variables	Sets
Measurement	Calibration
Dependent variables	Qualitative outcomes
Given populations	Constructed populations
Correlations	Set-theoretic relations
Correlation matrices	Truth tables
Net effects	Causal recipes

talk to their colleagues in the other camp. It soon became apparent, however, that my formalization of comparative inquiry offered more than a systematization of a specific type of macrosocial inquiry and also more than a window on important aspects of qualitative research. It provided as well the foundation for a new way of conducting social research, especially research focused on the study of cross-case patterns.

The conventional template for much social research, especially quantitative analysis of nonexperimental data, is straightforward. It starts with the specification of a dependent variable and the various independent variables thought to explain variation in the dependent variable. Ideally, these specifications are based on theory and existing research literatures. The preferred situation is one where the researcher can link different independent variables to competing theoretical perspectives. Researchers typically test their theories by collecting relevant data from taken-for-granted, given populations of observations, focusing either on the entire population, if the number of cases is small, or on a sample of observations, if the number is very large. The researcher then develops and tests one or more models. The focus of each test is typically on the relative explanatory power of competing independent variables. Explained variation in the dependent variable is partitioned according to the independent variables associated with different theoretical perspectives. Table A lists seven key elements of this template and the corresponding elements of the alternate template prompted by *The Comparative Method* and developments that followed from its publication.

VARIABLES VERSUS SETS. The lifeblood of the conventional template is the variable. A variable captures a dimension of variation, an aspect that varies by level, degree, or kind across cases. Variables sort, rank, or array cases rela-

tive to one another, often based on the arithmetic mean. For example, some countries are "more democratic"; some are "less democratic." Some individuals have "more income"; some have "less income." A set, by contrast, is a grouping and is more case-oriented than a variable because sets entail membership criteria and have classificatory consequences. While a variable can be labeled "degree of democracy," a set cannot, because this label does not group cases using membership criteria. However, it is possible to define and construct "the set of democratic countries" and to list the relevant members of this set.

This is not to say that a set is a mere nominal-scale classification, which is the usual preconception. Cases can vary in the degree to which they satisfy membership criteria, which is the inspiration behind fuzzy sets. With fuzzy sets, membership scores can range from 0.0 (nonmembership) to 1.0 (full membership). Fuzzy sets are simultaneously quantitative and qualitative. Full membership and full nonmembership are qualitative states; between these two are varying degrees of membership in the set, with 0.5 the qualitative crossover point between "more in" and "more out." The assignment of set membership scores follows directly from the definition and labeling of the set in question.

MEASUREMENT VERSUS CALIBRATION. Measurement in conventional social science is usually based on the use of indicators. Indicators must meet a minimum requirement, namely, they must array cases in a way that reflects the underlying construct. Cases' scores are evaluated relative to one another, based on inductively derived, sample-specific statistics such as the mean and standard deviation. For example, a "high" score is well above the mean score; a "low" score, well below. All variation in an indicator is usually treated as meaningful and taken at face value.

The conventional variable is uncalibrated. Cases' scores are understood only relative to one another. For example, it is possible to state that one country is more democratic than another or even more democratic than average but still not know if it is more a democracy or an autocracy. Calibration, by contrast, involves the interpretation of measures relative to external standards. We know, for example, that at 100°C water boils and at 0°C water freezes. The use of external standards makes it possible to calibrate measures. For example, to calibrate degree of membership in the set of "rich countries" (the target set) using an uncalibrated variable such as per capita GNP, it is necessary to specify the score that would qualify a country for full membership, the score that would completely exclude it from this set, and the qualitative

breakpoint separating "more in" from "more out." How a set is conceptualized and labeled is crucially important to its calibration.

DEPENDENT VARIABLE VERSUS QUALITATIVE OUTCOME. The dependent variable is often the focal point of the conventional template. Researchers conventionally assume that the goal of research is to explain cross-case and/or longitudinal variation in the chosen dependent variable. For example, following the conventional template, a researcher interested in "welfare state retrenchment" might attempt to explain variation in levels of welfare spending across the advanced industrial societies over the past several decades, focusing on the independent variables linked to changes in expenditure levels. Different theories offer different explanations of this variation.

In the alternate template, researchers first conceptualize the phenomenon to be explained as a qualitative outcome—an observable change or discontinuity. For example, instead of trying to explain variation in levels of welfare spending across countries and over decades, the researcher would conceptualize "welfare state retrenchment" as an outcome that has occurred in specific times and places. How do we know "welfare state retrenchment" when we see it? The researcher must specify its key features. After identifying several good instances and studying them in some depth, the researcher could then develop criteria for assessing the degree to which different cases (across time and space) have membership in this outcome. Using these criteria, it would be possible, in turn, to calibrate degree of membership in the outcome as a fuzzy set. An important key to assessing outcomes is the use of substantive external criteria to define the phenomenon of interest and to evaluate its degree of expression. In general, defining a qualitative outcome (e.g., welfare state retrenchment) and assessing the degree to which cases have membership in it requires more researcher input, much of it interpretive in nature, than simply selecting a dependent variable and trying to account for its variation (e.g., longitudinal and cross-case variation in welfare spending).

GIVEN VERSUS CONSTRUCTED POPULATIONS. Most applications of conventional methods use either given (i.e., taken for granted) or convenient populations or samples derived from such populations. The ideal-typic population for social research is a national random sample of households. Researchers are also fond of using given populations that are of interest to the state, to corporate actors, or to specific audiences. Such populations typically have face validity, and researchers rely on this to justify their use.

If researchers are interested in explaining qualitative outcomes, however, it is sometimes a mistake to use a given population. Research on qualitative outcomes typically begins by identifying good instances of the qualitative outcome in question. In-depth research on these cases helps to define and clarify the outcome and establish membership criteria. Once positive instances have been identified and examined, it is possible to construct the population of candidates for the outcome, embracing both the positive and the relevant negative cases. When researchers avoid given populations and instead construct their populations, negative cases are not known at the outset but can be identified only once their "candidacy" for the outcome has been established. Thus, the alternate template dictates careful selection of negative cases.

It is important to note one of the key hazards of using given populations. Often, given populations include irrelevant cases—those that are not plausible candidates for the outcome. Because such cases lack the outcome and usually register low scores on relevant causal conditions, they inflate correlations, making them much stronger than they would be if the irrelevant cases were excluded. In other words, irrelevant cases, paradoxically, can appear to be theory confirming when they are in fact simply irrelevant.

CORRELATIONS VERSUS SET RELATIONS. Conventional quantitative social science is based almost entirely on correlational analysis. From multiple regression to factor analysis to structural equation models, all that is required for these procedures is a matrix of bivariate correlations, along with the means and standard deviations of the variables included in the matrix. It is important to understand that the correlation coefficient is completely symmetric in its calculation. Thus, when correlation is used to assess the connection between the presence of a cause and the presence of an effect, it tests equally for a connection between the absence of the cause and the absence of the effect.

Because it is symmetric, correlation is blind to set-theoretic relationships. The most common set-theoretic relation is the subset relation. For example, the observation that the developed countries are democratic is set relational: the developed countries constitute a subset of the democratic countries. Unlike correlational relationships, set-theoretic relationships are asymmetric. The assertion that "the developed countries are democratic" does not require that the not-developed countries be not-democratic. There can be many not-developed countries that are democratic, and their existence does not count against the initial claim, which is asymmetric. More generally, the assessment

of both sufficiency (shared outcomes) and necessity (shared antecedents) is fundamentally set theoretic and asymmetric.

CORRELATION MATRICES VERSUS TRUTH TABLES. Most conventional quantitative methods parse matrices of bivariate correlations or their mathematical equivalents. These calculations, in turn, assess how well two series of values parallel each other across cases. Using correlations, there is no direct examination of how case aspects fit together within cases. While the calculation of the "net effects" of an "independent variable" may seem to take account of "other variables," the consideration of other variables is based on formulas that are nothing more than simple manipulations of bivariate relationships. There is no direct consideration of how case aspects combine in empirical cases or the consequences of these combinations.

A truth table, by contrast, is a direct examination of the kinds of cases that exist in a given set of data. It lists the different combinations of causally relevant conditions and treats each combination as a different "kind" of case. Cases with the same profile of causal conditions are grouped together, making it possible to assess whether they agree on the outcome. Each combination of causal conditions is examined on its own terms, as a specific set of circumstances. If the cases with a given combination of causal conditions disagree on the outcome, this is taken as a signal that other causal conditions should be added to the truth table or that the truth table needs to be respecified in some other way, based on the comparison of positive and negative cases in each contradictory row.

NET EFFECTS VERSUS CAUSAL RECIPES. The conventional template for social research emphasizes theory testing as a competition between variables to explain variation in a dependent variable. The calculation of the net effect of each independent variable and the partitioning of explained variation in the dependent variable are central tasks in this approach. The net effect and statistical significance of each causal variable are based on its unique (nonoverlapping) contribution to explained variation in the dependent variable. The greater an independent variable's correlation with the dependent variable and the lower its correlations with its competitors, the greater its net effect. Net-effect thinking isolates causal variables from one another and attempts to purify the estimate of each variable's separate effect.

In case-oriented research, by contrast, researchers often focus on how causal conditions combine to generate outcomes. The idea that causal conditions

have "independent" effects that can be estimated separately runs counter to this fundamentally recipe-like understanding of how social causation works. Furthermore, more than one causal recipe may generate the same outcome, and no single cause may be either necessary or sufficient. Because the focus of causal recipes is on how causal conditions combine in each case, the alternate template is much more case centered than its conventional counterpart.

SUMMARY. The alternate template sketched above brings innovative analytic strategies to the examination of cross-case evidence. It transforms cross-case analysis in ways that make it much more case oriented, which in turn enhances the dialogue between cross-case analysis and within-case analysis. It also makes explicit the interpretive nature of social research. Rather than relying on simplifying analytic assumptions (e.g., that each causal condition has an independent impact), the alternate template relies on the researcher's substantive and theoretical knowledge and interests. Finally, this new template also allows a much tighter connection between theory, which is mainly composed of verbal statements embodying asymmetric set relations, and empirical research, by making social research explicitly set theoretic.

LOOKING AHEAD

John Gerring (2001: 207; see also Gerring 2012) describes QCA as one of the few genuine methodological innovations of the past several decades. A Google Scholar search in 2013 reveals that *The Comparative Method* has been cited nearly four thousand times, making it one of the most cited methodological texts in the social sciences. Since its launch in 1987, QCA has been used in more than 750 studies resulting in publication.[2] At last count, QCA articles have appeared in more than 220 peer-reviewed journals. More important, the community of scholars who are actively involved in applying and developing QCA and set-theoretic methods continues to grow, resulting in wider diffusion, application, and critical reflection.

What about the next quarter century? To be sure, QCA will remain the subject of intense methodological debate (see, for example, Lieberson 2004; Ragin and Rihoux 2004a; Ragin and Rihoux 2004b; Seawright, 2004; Ragin 2005; Seawright 2005). However, it will also increasingly be one of the methodological tools in the toolbox of many researchers and will be applied in a growing variety of research projects, across the social sciences (Poteete, Janssen, and Ostrom 2010) and beyond. An increasing number of social scientists

have been opting for multiple-case studies as a research strategy. This choice is based on the need to gather in-depth evidence on cases, capturing their complexity, while laying the foundation for limited, cross-case generalizations. This development coincides with a renewed interest in case-oriented research (see Mahoney and Rueschemeyer 2003; Gerring 2004; George and Bennett 2005; Gerring 2007; Byrne and Ragin 2009) and in new attempts to engage a more productive interaction between the quantitative and qualitative traditions and to develop mixed-method research designs (Brady and Collier 2004; Moses, Rihoux, and Kittel 2005). *The Comparative Method* and related QCA publications have played a prominent role in these discussions, offering researchers a set of tools for the systematic comparison of cases as configurations.

Set-analytic social science is still in its infancy. *The Comparative Method* was but a first step on an important journey in social scientific inquiry. There has been considerable progress along this path since *The Comparative Method* was published, but there is still much work to be done. In some respects, set-analytic approaches can be seen as challenging conventional quantitative approaches. A more productive view sees set-analytic approaches as offering a means for bridging the quantitative-qualitative divide, providing a way for qualitative and quantitative forms of analysis to achieve greater complementarity. Much of this bridging comes in the form of mixed methods and related techniques for joining discourse on variables and discourse on cases. Another important part of the bridging effort, however, comes in the simple recognition that there are very basic differences between the two approaches that must be taken into account in any attempt to achieve analytic complementarity. There is a fundamental mathematical and conceptual difference, for example, between studying asymmetric set-relational connections and studying symmetric correlational connections. Ignoring this and other basic differences can only cause qualitative and quantitative researchers to continue to talk past one another, despite their deeply shared goal of understanding social phenomena.

NOTES

Some of the arguments sketched in this introduction are discussed in greater depth in Ragin 2013 and Marx, Rihoux, and Ragin 2013.

1. fsQCA software can be freely downloaded from www.fsqca.com.

2. The estimate of 750 publications is based on an extensive mapping exercise of a group of researchers affiliated with the COMPASSS (COMPArative Methods for Systematic cross-caSe analySis) network. COMPASSS is a broad network of scholars and practitioners involved in the development of systematic cross-case analysis, QCA in particular. This survey of publications was conducted by Priscilla Álamos, Damien Bol, Benoît Rihoux, and Alrik Thiem as part of a concerted COMPASSS effort. A full overview is available at www.compasss.org.