REVIEW ESSAYS

Peter Abell

FOUNDATIONS FOR A QUALITATIVE COMPARATIVE METHOD

RAGIN, CHARLES C. The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies. University of California Press, Berkeley, Los Angeles, London 1987. xvii, 185 pp.

Charles Ragin has given us a most important book. Written in a delightfully accessible style, it provides a systematic framework for the analysis of causal models where the pattern of causality is complex, where the number of cases (observations) is small (either because of difficulties in data gathering or because only a few cases naturally occur) and where the data is essentially qualitative (i.e. dichotomous). Ragin has, thus, laid the foundation for a qualitative comparative method. His book will prove to be essential reading for comparative historians, for qualitatively orientated social scientists and for anybody who cares about the balanced development of the social sciences.

Despite the intellectual pre-eminence of what Ragin calls the "variable-oriented approach", based, in its most modern variant, upon the generalized linear model, a great deal of socio-historical data proves entirely resistant to treatment in its terms. The requirement of large data sets, when causality is both multiple and complex, renders statistical tools inappropriate and attempts to force limited data sets to conform to the assumptions of standard models invariably leads to indecisive results.

"Comparative" social scientists (a term which is literally misused, since all social science is essentially comparative but which in practice indicates those social scientists who whilst adopting a comparative case study approach often wish to make generalizations) are often deeply ambivalent about statistical techniques. On the one hand, in recognition both of their undoubted analytical power and of their legitimating function in the eyes of a "scientifically" orientated intellectual culture, one sometimes witnesses an unseemly scramble to jump on the bandwagon. On the other hand, the

International Review of Social History, XXXIV (1989), pp. 103-109

inevitable contrivance of so doing is acknowledged, but there are, nevertheless, few yardsticks for good practice outside the framework of statistical inference. All agree that the rigours of the experimental method are rarely if ever appropriate, so how can the comparativist be reasonably assured, when inspecting a handful of cases, that s/he has got the (causal) story straight?

Comparativists have had little to guide them beyond Mill's dicta – that is, until now. I think it may not be an overstatement to suggest that Ragin has changed things. Though the techniques which he offers us have a well-established pedigree in formal (Boolean) logic and the theory of circuits (in electronic engineering), the sensitivity with which he has adapted them to the purposes of the comparativist will, I believe, be universally applauded.

The book may be conveniently considered as comprising of three sections. Chapters 1 to 5 which, firstly, set up the problem and point to the ubiquity of small-number comparative studies; secondly, compare the "Case-Oriented Comparative Method" with the "Variable-Oriented Approach"; and thirdly, suggest strategies for combining these latter two approaches. Chapters 6 and 7 – the meat of the analysis – which are relatively formal and might prove uneasy reading for those with little background in formal reasoning. Finally, Chapters 8 and 9 which present some applications of the Boolean technique. The content of the book is intellectually cumulative and consequently requires a fairly close reading from the beginning although, I suspect, those readers conversant with the variable oriented approach, could start at Chapter 6.

In the early chapters, after pointing out that there is much confusion abroad concerning the precise nature of the comparative method, Ragin argues that any adequate social-scientific explanation characteristically involves "multiple conjunctural causation". By this, he means that recurrent events will usually have a number of *alternative* causes, each of which is comprised of a number of *conjoined* events. For instance, to take an example from later in the book, "the success of strikes" may causally depend upon "a booming product market" *and* "a large strike fund" *or* "a low strike fund" *and* "a threat of sympathy strikes". If the theory is correct then either of these conjoined conditions is sufficient to guarantee the success of a strike. In practice, of course, linear additive models so often used within the variable orientated tradition postulate similar patterns, if interaction conditions are permitted to enter into each of the additive terms. For instance, in the above example, the variable "strike fund" would appear as a positive interaction in one term (i.e. with "a booming product market") and as a negative interaction in the other (with "a threat of sympathy strikes"). Attempts to estimate a variable-orientated model would, however, prove difficult as a high level of implied correlation

(multicolinearly) would be induced between the alternative causes, making it difficult to estimate their relative potency. So, even if a large number (N) of cases could be found, statistical estimation might well prove difficult.

Ragin breaks with statistical tradition in two senses: firstly, by enabling analysis with a low N and, secondly, by treating each case as of equal importance rather than as contributing to a weighted average, as statistical models invariably do. This is exciting, though the approach is, as a consequence, exposed and vulnerable, being crucially dependent upon the accurate measurement of each and every case. A misattribution of any one causal factor could often lead to very different results, whereas with statistical modelling, because of the averaging techniques, this is less likely to happen. Nevertheless, central to the low-N comparative case-study approach is the analyst's detailed acquaintance with the material and thus, hopefully, a low likelihood of misattributions. For instance, we are unlikely to find that Skocpol[1] has misdescribed the causal factors she adduces; but if she has, it could become a matter of open scholarly debate and the implications of alternative descriptions could easily be studied if Ragin's models were to be used to analyse her data.

So, what is the "Boolean" approach to Qualitative Comparison? Chapter 6 gives us the basic concepts and Chapter 7 some "extensions". Those familiar with Boolean algebra will anticipate the essentials of the technique, though those who recoil from the whole idea of "algebra" will no doubt be pleased to hear that very little knowledge of mathematical manipulation is needed to appreciate the way it works. Furthermore, the beauty of Boolean techniques is that they follow the every day logic of "off" and "on" (or true and false) so readers versed in basic propositional logic will feel quite at home.

Data from comparative qualitative case-studies may, Ragin avers, always be conveniently arranged as an array (or matrix) – where the rows represent each case, the columns the causal conditions/factors and one column the "outcome" which is to be causally explained. For example, and using letters for convenience, three causal factors might be proposed to account for "military regime failure" (F) – the outcome; there may be "conflict between older and younger military officers" (A); "Death of a powerful dictator" (B) and "CIA dissatisfaction with the regime" (C). The array would now have three columns for the "causes" and entries would be 1 if the factor was present, and 0 otherwise. The F column would, likewise, take values, 1 = "failure" and 0 = "no failure". We see here the qualitative nature of the technique, all the factors (causes and outcomes) are scored dichotomously. The Boolean approach permits us to analyse such arrays –

[1] Theda Skocpol, *States and Social Revolutions. A Comparative Analysis of France, Russia and China* (Cambridge, 1979).

however simple or complex – in a systematic way so that patterns of "multiple conjunctural causation" can be detected.

It is useful to see how this is done. Firstly, the array may be reduced by collapsing any identical rows and noting their frequency. As I mentioned before, the technique is not based upon considerations of relative frequency so a unique row is given the same weighting as one which repeatedly occurs. Each row of this reduced array now represents a particular constellation of causal variables and an outcome (assumed accurately described). One may then examine the rows which have a positive outcome (scored 1 on F in the above example). Each will necessarily have a unique pattern of 1 and 0 entries. Ragin adopts the convention of using upper case letters to denote the presence of a factor (score 1) and lower case to denote its absence (score 0). So, in the above example the data he gathers lead to an equation as follows:

$$F = Abc + aBc + abC + ABc + AbC + aBC + ABC,$$

where each additive term (e.g., Abc, aBc) represents one row in the reduced array. We may read this equation straightforwardly as: F is caused by either the conjunction of A, not B and not C or the conjunction not A, B, and not C or . . . etc. So, here we have seven empirically given patterns of "multiple conjunctural causation". Abc means we have at least one case of regime failure which had "conflict between older and younger military officers" only, because there was neither "death of a powerful dictator" nor "CIA dissatisfaction with the regime". The above equation for F (and note we still have made no mention of not-F i.e. regime continuation, which may well appear in the data) is not the most parsimonious causal account we can arrive at. It merely re-describes some of the rows (scored F = 1) in the reduced array. It is, nevertheless, a useful summary and with complex data arrays one should not underestimate the heuristic advantages of merely going this far. Be this as it may, the Boolean approach permits us to search for a more parsimonious causal pattern. Although Ragin does not quite set up the search in this manner, it can be conceived of as involving two steps. Firstly, a further matrix or array is constructed out of the complex terms in the equation (it is called the primitive sums-of-products equation) whereby we compare each alternative conjoined term with every other term (i.e. by paired comparisons). Now, clearly, if any pair of terms is identical except for *one* variable then that variable can be dropped – it is causally irrelevant. For example, Abc and ABc can be combined to give Ac. Repeated application of this procedure can much simplify the equation and in fact the one above reduces to:

$$F = A + B + C.$$

Now, of course, such clean reductions will not always occur. For example, the hypothetical equation:

F = AbC + aBc + ABc + ABC,

will reduce by paired comparison to:

F = AC + AB + BC.

This allows for the introduction of the second step. Again, a matrix may be constructed, this time comparing the terms in the final equation above (called prime implicants) with those in the preceding (original) equation. We find that any one of the implicants implies (in reduced form) several of the original terms. For instance, AC implies ABC and AbC, AB implies ABC and ABC and BC implies ABc and aBc. There is still redundancy here – the "final" equation above is not as causally parsimonious as it might be. Ac and Bc together are sufficient to generate F. Thus, our final reduced equation becomes:

F = AC + Bc.

This is precisely the pattern we encountered earlier in respect of the causes of successful "strikes". Here, either the causal factor A in the presence of C or the factor B, in the absence of C, is sufficient to produce F.

So far so good – complex arrays of absence/presence data can be reduced to the most parsimonious causal account. Furthermore, if the computations are beyond the reach of pencil and paper then computers will do the dirty work for us. But what of those cases which do not become F? Comparative method usually entails the juxtaposition of cases that are and are not F! In the variable orientated approach variance is developed upon the dependent variable. The above equations account only for the subset of cases (in the original array) which were Fs (i.e. regime failures). We could repeat the whole exercise for the not-Fs but in practice (if the data is consistent – see below) there is no need for this. Ragin shows how we can use De Morgan's law to generate the most parsimonious causal account of the cases which do not become F. This is merely done by switching "and" for "or" and "or" for "and" in the equations for F. So if as before:

F = Ac + Bc,

then

not-F = f = ab +aC +bc.

Empirical data might, however, not come this clean. Ragin offers us a technique for dealing with "contradictions". Suppose we find cases where Ac leads to F and to f (i.e. not-F) . . . what can we do? The obvious reaction

https://doi.org/10.1017/S002085900000907X Published online by Cambridge University Press

to such an occurrence is to postulate a further factor which conjunctively discriminates (AcE and Ace) between the cases and then to gather the appropriate data. In fact, "contradictions" will always be indicative of greater causal complexity as long as we are convinced that our established descriptions (coding) are accurate. They thus provide, under this assumption, the occasion for a further elaborated theory. Ragin also proposes a number of other Boolean strategies when for one reason or another additional data are not available; possibilities include coding contradictory cases as either zero or unity. He briefly explores the virtues of both these possible research strategies and even, when the number of cases warrant it, advocates coding the outcome in terms of a threshold probability. Whether or not such strategies will prove useful only research experience will, I suspect, demonstrate. What is certain though is that he has opened up a much needed line of communication between variable orientated and case study research.

The original array of empirically derived data may or may not contain examples of all the logically possible combinations of causal factors. Something can be made of this – the problem of "limited diversity" as Ragin calls it. Clearly, further cases might, when examined, reveal additional causal patterns or, alternatively, they may all conform to those already encountered. Indeed, the Boolean technique can be used to locate empirical typologies of "existing" as opposed to "non-existing" combinations of causes. These must always be provisional. Further research might locate "non-existing" combinations but a detailed inspection of the particular constellations which persistently fail to occur can be very revealing as Ragin shows in his closing chapters by reanalysing data from Stapleton *et al.*[2]

Received theories may propose alternative patterns of causal determination; empirical research may provide evidence for yet further patterns. The Boolean approach can be used to assess, in a precise way, the degree to which theories are or are not confirmed by the evidence. This may prove to be the most significant contribution of the Boolean technique. Since the patterns of disjunctive conjunctions can be treated as sets which may intersect, theoretically and empirically derived patterns of causality may be studied for their intersection, complements, union etc. Such procedures allow us to answer questions like "What are the major shortcomings of existing theories?" The interested reader will have to study Ragin's text for details but as the author says, he only scratches the surface of a potentially very rewarding line of inquiry.

In Ragin's capable hands, the Boolean approach proves to be an eminently flexible instrument specifically designed for detecting the impact of

[2] Vaughn Stapleton, David Aday and Jeanne Ito, "An empirical typology of American metropolitan juvenile courts", *American Journal of Sociology*, 88 (1982), pp. 549-564.

patterns of multiple conjunctural causation upon a specific outcome (or dependent dichotomous variable). But does this go far enough? If nothing else, the variable centred approach has taught us, firstly, that spurious correlation between variables is endemic (i.e. the correlation of two variables is not causal but attributable to a common causal antecedent) and, secondly, that causal predictor variables are often causally related amongst themselves. In the jargon of statistical modelling we have to move from single-equation to simultaneous-equation models, if we are going to do justice to the real world. Can Ragin's small-N comparative approach accomplish something parallel to this? Although he does not take us in these directions, it does seem that his methods are sufficiently flexible to presage some hope that the answer is in the affirmative.

First, spurious correlation. Suppose we examine the original array without depicting a particular outcome column and find that two columns are either identically or opposingly coded, then in either case, there is a perfect correlation between them (positive in the first case, negative in the second). We could then treat the *pair* of columns as the outcome, score them either as unity or zero and use the standard reduction techniques to search for alternative common causal antecedents.

Second, simultaneity effects; things here are a little murkier but not beyond hope. Suppose, rather than saying X or Y cause F, we want to say X causes Y and Y causes F. In the former case we can obtain observations of the form: $F = 1, Y = 1$, and $X = 1$; $F = 1, Y = 1$ and $X = 0$; and $F = 1, Y = 0$ and $X = 1$. These reduce to:

$$F = X + Y.$$

In the latter case, however, only $F = 1$, $Y = 1$ and $X = 1$ can occur. This pattern is of course compatible with F being caused by X *and* Y but here $F = 0, X = 1$ and $Y = 0$; and $F = 0, X = 0$ and $Y = 1$ can occur. The three models can thus be distinguished by noting not only the cases where $F = 0$ and $F = 1$ but combinations of the causal factors that do not occur as well. At the moment I see no way of distinguishing, though, between a model where X causes Y and Y causes X, on the one hand, and where this is still true though X has in addition a direct effect on F. Nevertheless, I have managed to convince myself that rather complex multiple equation models can be handled using Ragin's techniques.

In conclusion then, Charles Ragin has not only given us the logical foundations for a systematic small-N comparative method, he has done so in a manner that persistently provokes further thoughts and ruminations. His book is a joy to read; it has the characteristics of all truly creative works, it starts with a simple but fertile idea and coaxes from there within a whole new vision.