The Rationality of Drawing Big Conclusions Based on Small Samples: In Defense of Mill's Methods*

JUKKA SAVOLAINEN, *State University of New York at Albany*

Abstract

Skocpol endorses the application of Mill's methods of causal inference for comparative historical explanations. According to Lieberson (1991), in studies where the sample size is very small, Mill's methods are inappropriate because they: (1) do not allow for probabilistic theories; (2) cannot handle interaction effects; (3) cannot accommodate multiple causes; (4) require the absence of measurement errors. Each of these claims turn out to be incorrect due to confusion over the uses of Mill's methods, failure to appreciate the aims of case-oriented explanations, and a narrow conception of cause. Small sample size does not constitute an obstacle to the application of Mill's methods.

Skocpol has advocated the use of Mill's methods of *agreement* and *difference* as appropriate logics of causal inference for comparative historical research (Skocpol 1979, 1984; Skocpol & Somers 1980). In 1986, a debate took place between Elizabeth Nichols and Skocpol concerning the value of Mill. Recently, in the pages of this journal, Stanley Lieberson (1991) launched a new attack on Skocpol's methodological agenda.[1] According to Lieberson, causal inference operating within this framework is committed to assumptions that violate prevailing standards of sociological inquiry. He argues that "application of Mill's methods to small-N situations does not allow for probabilistic theories, interaction effects, measurement errors, or even the presence of more than one cause" (318).

In what follows, I will challenge each of these four claims in turn. Let me begin, however, by clarifying what it actually means to use Mill's methods in this context. This is necessary, since Lieberson is confused over the fact that

Skocpol only endorses the use of Mill's methods as rules to eliminate incompatible causal claims, not as providing a logic of discovery or a canon of proof.

The Uses of Mill's Methods

According to Mill, his methods of difference and agreement enable to *discover* and *prove* causal relations. Ever since Cohen and Nagel's (1934) discussion of Mill's methods it has been a philosophical commonplace that they fulfill neither of those functions. Instead, the value of Mill's methods is in their capacity to *eliminate* a limited set of alternative causal statements. The following is Mill's formulation of the method of agreement:

If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agrees is the cause (or effect) of the given phenomenon. (Mill, cited in Cohen & Nagel 1934:251)

To illustrate the limits of this rule, consider Table 1. It follows from Mill's reasoning that $C1$ is the cause of P, because only that circumstance is shared by both instances of P's occurrence. Now let us apply this scheme to the explanation of revolutions. Let instance 1 denote the French Revolution and instance 2 the Russian revolution. How are we to proceed from here, according to Mill, in order to discover the causes of revolutions? In practice, we cannot. In order to put the method of agreement to any use it is mandatory to restrict the number of possible causes to those we consider as relevant candidates. The number of "circumstances" characterizing prerevolution France and Russia is infinite. Since this method fails to inform us what to select from that pool of potential causes, it cannot be regarded as a method of discovery. Furthermore, since it is always possible to construct explanations drawing on those circumstances that were absent in the previous applications of the method, it can neither prove any claims.

Mill's method of agreement is, then, worthless as a method of discovery and fallacious as a canon of proof. Its true value is in its function to eliminate alternative explanations. The tenable import of the method of agreement can be formulated as follows: No factor can explain an outcome satisfactorily that is not common to all occurrences of that outcome (Cohen 1989:260).

This analysis applies equally to the method of difference, which, according to its contemporary interpretation, asserts that no factor can explain both an outcome and its opposite. This is but a complement of the method of agreement.[2]

It is obvious that when Skocpol (1986) promotes the use of Mill's methods in comparative historical research, she only has the eliminative function in mind:

In the research for States and Social Revolutions I did exactly that. Using Mill's logics, I considered carefully whether causal hypotheses from Marxian class analysis and from Ted Gurr, Neil Smelser, Charles Tilly, and Chalmers Johnson could do an adequate job in explaining why France, Russia, and China had social revolutions, while other similarly situated modernizing agrarian states did not. (189)

TABLE 1: The Method of Agreement

	Circumstances	Phenomenon that occurs	Cause of P
Instance 1	C1, C2, C3	P	
			C1
Instance 2	C1, C4, C5	P	

Skocpol and Somers (1980:186) mention Robert Brenner's article *Agrarian Class Structure and Economic Development in Pre-Industrial Europe* (1976) as an exemplar of the application of the method of difference, because, according to them, Brenner's article "employ(s) comparative history to refute alternative, competing arguments." When considering Lieberson's criticisms it is imperative to have established that Skocpol does not invest Mill's methods with powers of discovery or proof.

Case-Oriented Explanations and Probabilistic Theories

According to Lieberson (1991), "application of Mill's methods to small-N situations does not allow for probabilistic theories" (318). If this were true, it would mean serious limitations to the use of Mill's methods, for — as Lieberson points out — processes involved in macrosocietal outcomes are often probabilistic in nature.

Much of comparative historical sociology follows what Ragin (1987) calls the case-oriented approach. The goal of case-oriented research is to explain particular outcomes. The cases under study, such as social revolutions or the development of capitalism, are considered important in their own right. The comparative historical studies using Mill's methods fall into this category (Ragin 1987). The explanatory task in case-oriented approach is to identify the causal processes that brought about the outcome, the occurrence of which has been defined certain at the outset. In this sense these *explanations* can be perhaps understood as "deterministic." However, this framework does not require the assumption that the outcome *had to* happen, i.e., was somehow inevitable. Quite to the contrary, particularistic explanations can very well make use of probabilistic *theories* to aid historical interpretation. Using a small number of cases makes no difference in this respect: it is quite advisable to resort to stochastic principles when explaining an outcome of a single flip of a coin.

Consider Weber's (1958) account of the reasons why capitalism developed in western Europe, not in China — a study that Lieberson (1991:308), quite rightly, regards as an instance of the application of Mill's method of difference. Although Weber is not explicit about the issue, there is every reason to suppose that the processes that his explanation relies on are probabilistic in nature. His explanation would not suffer at all should some Protestants deviate from the expected pattern; neither does it weaken the argument to allow for the

occurrence of Catholic entrepreneurs investing at the same rate as their Protestant peers. Only the aggregated result of these individual behaviors is of importance.

To illustrate the inadequacies of "the deterministic model," Lieberson (1991) draws an analogy based on the following airline customer's dilemma:

Suppose a rude employee is encountered, luggage is lost, or the plane is delayed . . . Based on a small number of experiences, one may decide to shun a certain airline . . . However, conclusions drawn on the basis of such practices are often wrong . . . because small number of cases is an inadequate basis for generalizing about the process under study. (310-11).

Lieberson is, of course, perfectly right in what he says, but the analogy does not extend to case-oriented studies, the goal of which is not to provide generalizations beyond the scope of the study. This is what Skocpol and Somers (1980) have to say about generalizing:

Comparative-historical causal arguments cannot be readily generalized beyond the cases actually discussed. In the preface to Social Origins [of Democracy and Dictatorship], Barrington Moore likens the generalization his study establishes to "a large scale map of an extended terrain, such as an airplane might use in crossing a continent." This is an appropriate metaphor. And the reflection it inspires in this context is that no matter how good the map were of, say, North America, the pilot could not use the same map to fly over other continents. (195)

The aviation theme is the only shared element between this analogy and that of Lieberson. Indeed, the very aim of case-oriented research renders it suboptimal to producing generalizations. Comprehensive interest in specific historical outcomes drives attention to particularities.[3]

Handling Interaction Effects

According to Lieberson (1991), Mill's methods are unable to deal with interaction effects, and must therefore assume their absence. This claim, if valid, would render the range of application of Millian causal inference unacceptably narrow. Lieberson develops his argument drawing on hypothetical data about car accidents (Table 2).

The sample size of this setting is 2. In case 1, an automobile accident has occurred. The purpose of the study is to explain why the accident took place by comparing it to a case where it did not. The application of the method of difference leads to conclude that the accident (Y) was caused by car entering from right ($X2$). "We would also conclude," Lieberson (1991) argues,

that the accident is not caused by drunk driving or running of a red light because the variables are the same for both drivers yet only one had an accident. Such conclusions are reached only by making a very demanding assumption that is rarely examined. The method's logic assumes no interaction effects are operating (i.e., that the influence of each independent variable on Y is unaffected by the level of some other independent variable.) The procedure cannot deal with interaction effects; it cannot distinguish between the influence of inebriation or running a red light from another constant, such as the benign fact that both drivers were not exceeding the speed limit. (312)

TABLE 2: Application of the Method of Difference[a]

	Accident (Y)	Drunk Driving (X1)	Car Entering from Right (X2)	Driver Speeding (X3)	Runs a Red Light (X4)
Case 1	+	+	+	-	+
Case 2	-	+	-	-	+

[a] Source: Lieberson 1991:313

Lieberson seems to entertain an exceptionally restricted ontology of cause. To him a "cause" denotes a single variable that, due to its pragmatically indivisible nature, I will call an *atom variable*. Yet Mill's methods provide no constraints regarding the formal composition of the entities one chooses to treat as causes. Those may very well feature structures of highest complexity; an interaction between two variables is by no means ruled out as a causal candidate.

To witness that Mill's methods not only allow for interaction effects, but that their application in actual research has resulted in the isolation of such, one needs to look no further than Skocpol's *States and Social Revolutions* (1979). Ironically, another critique of Skocpol's methodology, Elizabeth Nichols (1986), accuses Skocpol of misusing Mill precisely *because* she analyzes a combination of variables rather than single (atom) variables. This gives Skocpol (1986) an opportunity to be explicit about the interactive nature of her causal argument:

> I show that state breakdown and peasant revolts both occurred in France, Russia, and China, but that in England and Japan there were state breakdowns without peasant revolts, while in Germany 1848 the state breakdown was only temporary and peasant revolts were regionally limited, and in Russia 1905 the state breakdown was very temporary and soon reversed. This, I submit, is a very powerful way to establish that *both* breakdowns of monarchical state machineries *and* peasant revolts, taken together, spurred social revolutionary transformations in France, Russia, and China. (189, emphasis added)

Small-N studies using Mill's methods have thus no problem in accommodating interaction terms. Lieberson is right, however, in saying that "the method cannot consider the possibility that there are interaction effects between two variables" (Lieberson 1991:318). The *method* cannot consider that, because it is no method of discovery. It is up to the research community to put forward hypotheses, whether they involve interaction terms or not, and to design the studies accordingly. If one is interested in the interactive effect on car accidents of, say, drunk driving and running against a red light, one should design the setting differently from the one displayed in Table 2.

On Measurement Errors

Lieberson claims that the application of Mill's methods to small samples must assume error free measurement of theoretical concepts. This is possibly the most radical of Lieberson's accusations since there is no way any study could meet the requirement of perfect measures. As to the reasons why Mill's methods should necessitate such an assumption the reader is left in the dark. It is even unclear whether the measurement error argument is independent of the already refuted allegation that the application of Mill to small samples does not allow for probabilistic theorizing (Lieberson introduces the measurement error problem in the context of arguing in favor of a "probabilistic approach"). To the extent that the former presumes the latter, nothing remains to be added concerning the measurement error claim. In any case, the very problem of measurement error can be dismissed as artificial in the context of predominantly qualitative historical research, in which conceptualization and "measurement" are practically inseparable processes, moving in a hermeneutic circle.

Multiple Causes

Lieberson's last claim asserts that the application of Mill necessitates the unrealistic assumption that an outcome can only be a function of one cause. There are two aspects to this claim. First, to quote Lieberson: "Mill's method cannot work when more than one causal variable is a determinant and there are a small number of cases" (Lieberson 1991:314). This is tantamount to saying that Mill's method cannot handle multivariate additive model types of causal statements. Keeping in mind the preceding discussion concerning interaction effects it is obvious why this view is distorted. Lieberson understands by "cause" what was previously characterized as an atom variable — a conception to which the use of Mill does not require one to adhere.

Second, Lieberson may also mean, although this is not clear, that Mill's method of *agreement* is defective in that it does not take into account that the same outcome (car accident) may result from two or more *alternative* causes (drunk driving, or running against red light, or speeding, etc.). A classic treatment of this objection to using Mill is found, once again, in Cohen and Nagel (1934:269-72), of which the following is but a recapitulation.

A car accident can clearly result from more than just one chain of events. Does it follow from this that it is irrational to try to infer its cause based on just one instance? Insurance companies do that all the time, and believe to be successful in it too. The multiple cause argument is deceptive because it is based on a lack of symmetry in the analysis of causes and effects: not only are there many causes of car accidents, there are also different types of accidents; and, what is of importance here, there is a correlation between the two: a car accident looks different *depending* on the sequence of events that preceded it. As Cohen and Nagel (1934, 270) conclude: "When a plurality of causes is asserted for an effect, the *effect* is not analyzed very carefully. Instances which have significant differences are taken to illustrate the *same effect*."

Conclusion and Discussion

Contrary to what Lieberson claims, the application of Mill's methods to small-N situations *does* allow for probabilistic theories, interaction effects, measurement errors, and even the presence of more than one causal variable. Lieberson's discussion is ridden with serious misunderstandings. To restate the most obvious ones: confusion over the uses of Mill's methods, failure to appreciate the aims of case-oriented explanations; and adherence to an inappropriate notion of cause.

Not only do Lieberson's *criticisms* of Skocpol's methodology miss the target; consider the following cites taken from the *constructive* part of his text:

[Mill's] methods require . . . confidence that all possible causes are measured. (Lieberson 1991:315)

Because of the small N's and the reasoning Mill's methods require, it is vital to include all possible causal variables. (Lieberson 1991:317)

By suggesting that it is even feasible to consider all the possible causes of an outcome in a single study, Lieberson instantiates the following observation by Bernard Cohen (1989): "While critics [in sociology] attack the use of the [experimental] model, they invest the experimental method with magical powers in those circumstances where they believe it is applicable, powers of empirical proof" (249). Although there is no excuse for Lieberson to commit this error, it makes sense that he should. First, remember that he treats causes as atom variables. Second, he has strong background in quantitative survey sociology, in which empirical analysis is confined to a set of variables fixed by the data matrix. In this framework the number of "causes" (understood as variables) is indeed limited.

Notes

1. Since its appearance in *Social Forces*, a version of Lieberson's article was published in *What is a Case? Exploring the Foundations of Social Inquiry*, edited by Charles Ragin and Howard Becker (1992). Also, in a recent issue of *American Sociological Review*, Hooks (1993) cites Lieberson when discussing the limitations of small-N historical data in making "absolute conclusions" (38). Lieberson's views seem to be influential, and must therefore be taken seriously.

2. An interested reader should consult Cohen and Nagel (1934) or Cohen (1989) for details.

3. This is not to say that case-oriented research cannot yield results that have general import. On the contrary, they can, and often do, generate hypotheses about phenomena outside the realm of the cases under study. Skocpol's (1982) extension of her model to the revolution of Iran is a case in point.

References

Brenner, Robert. 1976. "Agrarian Class Structure and Economic Development in Pre-Industrial Europe." *Past and Present* 70:30-75.

Cohen, Bernard P. 1989. *Developing Sociological Knowledge. Theory and Method.* 2d ed. Nelson-Hall.

Cohen, Morris, and Ernest Nagel. 1934 *An Introduction to Logic and Scientific Method.* Harcourt, Brace.

Hooks, Gregory. 1993. "The Weakness of Strong Theories: The U.S. State's Dominance of the World War II Investment Process." *American Sociological Review* 58:37-53.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:307-20.

Nichols, Elizabeth. 1986. "Skocpol on Revolution: Comparative Analysis vs. Historical Conjecture." *Comparative Social Research* 9:163-86. JAI Press.

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* University of California Press

Ragin, Charles, and Howard Becker (eds.). 1992. *What is a Case? Exploring the Foundations of Social Inquiry.* Cambridge University Press.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China.* Cambridge University Press.

_____. 1982. "Rentier State and Shi'a Islam in the Iranian Revolution." *Theory and Society* 11:265-300.

_____ (ed.). 1984. *Vision and Method in Historical Sociology.* Cambridge University Press.

_____. 1986. "Analyzing Causal Configurations in History. A Rejoinder to Nichols." *Comparative Social Research* 9:187-94. JAI Press.

Skocpol, Theda, and Margaret Somers. 1980. "The Uses of Comparative History in Macrosocial Inquiry." *Comparative Studies in Society and History* 22:174-97.

Weber, Max. 1958. *The Protestant Ethic and the Spirit of Capitalism.* Scribner's Sons.