Edited by Vaughn R. McKim and Stephen P. Turner

# Causality in Crisis?

## Statistical Methods and the Search for Causal Knowledge in the Social Sciences

Clark Glymour

Peter Spirtes

Richard Scheines

Nancy Cartwright

James Woodward

David A. Freedman

Mary S. Morgan

Stanley Lieberson

# The Big Broad Issues in Society and Social History

## Application of a Probabilistic Perspective

*Stanley Lieberson*

When an important event occurs, we want to understand it in the sense that we want to know what causes it. This sentence involves questions that I will not try to define in an elaborate way. By "important" I simply mean an event that the larger society (or some subset) defines as such. It can also be "important" if it contributes to social science because of the question's theoretical relevance, such as linking up to empirical issues that have not been resolved. By "understand" I have in mind gaining a greater sense of the event in a variety of possible ways such as what the experience is like for those involved, the attitudes of society towards it, institutional and societal predecessors and/or correlates, the strategies and tactics used by the individuals and groups involved, the historical background, and the like. One specific kind of understanding involves "causality," which I shall view as determining the condition(s) which affect the likelihood of occurrence of the observed event.

My concern here is not about elaborate definitions of these terms. I will not focus primarily on the difficulties that a discipline such as sociology encounters when applying routine quantitative procedures to problems that generally involve a relatively large number of cases. Typical examples of this are studies of the factors influencing internal migration, unemployment, marriage, divorce, intergenerational occupational mobility, infant mortality, prejudices and other attitudes. Answering these questions is difficult enough; there are often serious questions about the logic of social research even under these relatively ideal circumstances (as other essays in this volume make clear). I will not delve into these matters other than to mention in passing that we often try to pursue experimental analogies based on nonexperimental data. Since it is difficult enough to do first-class experiments, just visualize the added difficulties when working with nonexperimental data. This is the typical situation in such primarily nonexperimental social sciences as sociology, economics, political science, and social anthropology, and it leads to a wide variety of difficulties. For example, there are extremely sticky difficulties in creating suitable controls, taking into account unmeasured selectivity, finding suitable comparisons that are themselves unaffected by the independent variable, or dealing with asymmetrical causal processes (for a more elaborate discussion of these issues, see, for example, Lieberson 1985; Singer and Marini 1987; Freedman 1991, and the literature cited therein).

I will start here with a different set of issues: namely, important kinds of events that, for one reason or another, cannot be studied through the analysis of large numbers of cases. As we shall see, there are special weaknesses in the methods that are customarily used for analysis in such instances. In considering these weaknesses, I will develop a probabilistic perspective that is relevant for analyses of social processes—regardless of data size. It is a perspective which has implications that I believe are not fully appreciated even when researchers work with large data sets. Some of the implications will be illustrated with counterfactual examples drawn from sports. From these examples, I then draw three major conclusions which—if valid—are applicable to all kinds of social science problems, regardless of the type of data. These involve a perspective on how social processes are best conceptualized given the limitations of our methodological tools (what I shall call a "multilayered probabilistic approach"), a discussion of the role of chain models in understanding events, and—returning to our initial question—how we might best approach the task of causally explaining specific events.

## Small N's When There Are Just Not Enough Cases for a Statistical Analysis

There are many questions about society that do not readily lend themselves to straightforward quantitative studies. Often only a small number of cases exist in the entire universe. (Remember that "small" and "large" are not fixed, because the minimum necessary size will vary by the number of variables, the complexity of the models, the strength of the relationships, and the like.)[1] In some problems, even if the total number of cases might be adequate, sometimes there are only a few that we know enough about to permit sufficient description and measurement. In other problems, the amount of effort necessary to research even one case is massive and it is impractical to expect researchers to go beyond one or two cases (as might well be the problem when intensive field work over the course of several years is necessary as in some sociological and anthropological studies). Or years in the library may be necessary to research an enormous number of historical documents to adequately study one case, and a sufficient number of cases for a statistical analysis may be a biological impossibility under present life expectancy for a single researcher. In any case, for one reason or another, the usual large-N statistical analyses are not easy for such questions as, for example: the consequences a specific social policy has for society; the reason a given piece of legislation is passed; the causes of a given war; the causes of revolutions; the decline of empires; the reason two nations differ on a specific social policy; the origins of capitalism; the political consequences of presidential assassinations in the United States; the causes of current black-white differences in income; the reason why there is an exceptional level of violence in the United States; the rising tide of nationalism in Europe; the economic success of several new nations in South East Asia; or the absence of economic progress in most African nations.



Incidentally, the social organization of research is not a trivial consideration here. What keeps a team of researchers, say a group of anthropologists, from each studying a different community with the goal of pooling the results at the end and doing a statistical analysis? Perhaps antipathy towards statistical analysis is itself a sufficient reason. In any case the cross-cultural files are constructed only as each individual researcher does his or her study.

In any case, we are considering problems about society which do not lend themselves to the customary statistical analyses. Under such circumstances, there are typically three approaches taken. One is to rely on a descriptive case study of one or two such events, with the notion that such a deep analysis will permit one to ferret out the causes of the events. (Although such work is sometimes labeled as descriptive, involving no explicit appeals to the notion of causality, I think the distinction may be less meaningful in practice because the "describer" in such a work makes a judgment about what circumstances leading up to an event are relevant or important enough to be emphasized and discussed. Likewise, analyses of the consequences of an event implicitly involve causal notions.) Incidentally, a case study of this sort can be quantitative and include complicated statistical models. A second "solution" to the problem is what I would call an artificial forced variance approach. The third approach involves what *appears* to be an application of the logic of causal inference described by John Stuart Mill in his *System of Logic* (1872).

All three face serious problems. The single case study is a powerful instrument for deeply delving into a given situation and can be extraordinarily illuminating and suggestive. But it is usually very difficult to truly sort out and isolate the causal factor(s). Most often, there are almost infinite possibilities for what is accounting for the observed relationship. Moreover, it is difficult to have much confidence about the generalizability of the results. Is this the only cause of the dependent variable? Is it the most common cause? Is it a relatively rare cause? All of these questions are difficult to answer with the case study. Incidentally, a study based on a large N, with all sorts of variables and elaborate statistical models, can still be viewed as a small-N study if it deals with only one unit but means to generalize beyond that. As my colleague John Campbell observed in a discussion of these issues, if there is a one-country analysis of taxes over time, then it is a study of one case (even if there are elaborate data sets for a long period of time) if it is meant to generalize about tax policies in other nations.[2]

The use of forced variance is not an uncommon approach because it allows the researcher to use standard quantitative procedures when facing a situation with either a small number of cases in the universe or a small number of available cases. By "forced variance" I mean a situation where the research problem is modified so as to increase the number of measurable instances. In point of fact, it normally involves creating a different problem (even when the variables are the same) because it pursues analysis at a different level than that at which the original problem was formulated. An example is described in Lieberson (1985, 110–14). The United States has an exceptionally high homicide rate when compared to nations with comparable levels of urbanization and industrialization. Why is this? It is suggested that racial differences in socioeconomic position cause criminal violence.[3] Based on data gathered for 125 large metropolitan areas in the United States, it is found that indeed there is a significant correlation between the level of violent crime in the metropolitan areas and the level of socioeconomic differences between the races. The study, you will notice, has shifted from an issue about differences *between* nations to an issue about differences *within* a nation. The variables are the same ones we would use if there were cross-national data: some indicator of violent crime and some indicator of racial gaps in socioeconomic circumstances. But we have dropped to a different level of analysis—a lower one which could hold but not be applicable to what occurs cross-nationally. The nation effect, per se, is dropped. I will not go into details about the assumptions that this step involves since they are detailed in Lieberson (1985, chapter 5). The point here is that one cannot blithely drum up some added variance by dropping down a level of analysis. It is not interchangeable with the evidence that would occur at the appropriate plane.

The third method, which involves applying the method of agreement and the method of difference proposed by Mill (1872), has been used in the area of comparative and historical research when the number of cases are limited (see the illustrations provided in Lieberson 1991). The method is appealing since it allows one to logically eliminate possible causes if they do not correctly account for all of the small number of cases under examination. It is also appealing because there are no shades of gray in evaluating evidence: either the theory holds or it fails. In other words, all of the evidence is assumed to be strong evidence—indeed, exceptionally strong—of a magnitude rarely encountered in social science. As a matter of fact, the method never allows for multiple causes of a given event.[4] Multiple causes are indistinguishable from non-causes. If more than one cause operates in a set of situations, then whether the method of agreement or the method of difference is applied, the causes will be rejected because they will be indistinguishable from variables which have no influence on the outcome.



Unfortunately, as Mill himself observed in a section of his book that seems to be ignored by those who insist on using these procedures, the method of difference and the method of agreement are not applicable to the kinds of problems encountered in the social sciences (see Mill 1872, Book VI, chapter VII, 469–78). If the operation of multiple causes is indistinguishable from non-causes, for example, then the method cannot work in the typical social science circumstance in which nonexperimental data involve more than one causal factor affecting the dependent variable. If all variables present in two or more cases are the same except for one, it is almost certain—as Mill argues—that there has to be a complex set of reasons underlying the fact that two societies, say, are identical on all counts but this one variable (a circumstance which these methods cannot address). I will not review the criticisms I have raised elsewhere (Lieberson 1991) or the excellent and related comments in an earlier paper by Nichols (1986). But there is one issue that should be noted in passing here. A central weakness of attempts to apply Mill's methods—in spite of his objections—is the inability of the method of difference or the method of agreement to yield probabilistic causal models. As a consequence, the methods are obliged to use deterministic assumptions about social processes even when they are patently inappropriate (Lieberson 1991, 309–12).[5] Also, causal inferences based on small data sets have difficulty guarding against conclusions reflecting random combinations of variables. Typically, a small number of cases are used to evaluate the influence of several possible causal variables on a given explanandum. The information on each variable is usually presented in a dichotomous form (e.g., present-absent in each cell, or high-low). If we visualize the score for each dichotomy being randomly assigned to each cell in such a table, there is a good chance that some "meaningful" conclusion will be obtained under this methodology such that only one variable is found to account for the outcome. This is particularly the case when variables are combined if they do not seem to hold individually. The structure of the analysis allows for meaningless conclusions to be generated.

In short, to my knowledge the three major procedures for dealing with small-N situations are deeply limited, at least as presently practiced. The assumptions used to infer causality in small-N situations are often unrealistic, but are driven by the fact that small-N situations are not suited to a probabilistic approach. It is a case of the perspective being driven by the data. Indeed, the all-or-nothing quality of causal ascriptions on Mill's methods virtually precludes an ability to identify causal factors at work in large or small data sets dealing with social events.

## A Probabilistic Perspective

There are a variety of reasons for taking a probabilistic approach even if we think in deterministic terms. Bear in mind that a deterministic perspective refers to a view that a given factor, when present, will lead to a specific outcome. By contrast, a probabilistic perspective is more modest in its causal claim, positing that a given factor, when present, will increase the likelihood of a specified outcome. When we say, "If $X_1$, then $Y$," we are making a deterministic statement. When we say, "The presence of $X_1$ increases the likelihood or frequency of $Y$," we are making a probabilistic statement. Obviously, if given the choice, deterministic statements are more appealing. They are cleaner, simpler, and more easily disproved than probabilistic ones. One negative case ($Y$'s absence in the presence of $X_1$) would quickly eliminate a deterministic statement (Lieberson 1991, 309).

The subjects studied in the social sciences often involve substantial measurement errors that are not easily avoided and are often of a non-random nature. Complex multivariate patterns are often present in naturally occurring data and these are likely to make a deterministic approach unproductive. In addition, we are usually unable to measure all of the variables that we believe are actual (or possible) influences on the problem at hand. Finally, there is the role of what we might call "chance" in affecting outcomes.[6] (See Lieberson 1991, 309–10 for a more elaborate discussion of how these factors work against using a deterministic model.) Here "chance" or "random" can be taken as referring to idiosyncratic factors at work that we cannot deal with, either because we do not know them or because there are so many that we could not manage them all without, in effect, trying to explain the history of the universe. In a certain practical sense, it is almost irrelevant whether they could all be taken into account given sufficient energy and effort—although we will see below that there is reason to suspect that they cannot be.

However, there are two additional reasons for comparing the probabilistic and deterministic approaches. First, explanations of particular events, even where the research driving an explanation is itself probabilistic and based on very large N's, are often formulated in an inappropriate deterministic manner. Second, a probabilistic approach suggests a different way of conceiving social processes and of theorizing about them that is appropriate for both large- and small-N work, for both what passes as quantitative and qualitative research.



## A Sports Example

The path leading to a championship in a competitive sport provides a convenient opportunity to consider the role of "chance" factors and is a relevant model for thinking about social events more generally. From a deterministic perspective, the champion team is the best team. If we could visualize a counterfactual condition such that we roll time back to the beginning of the entire season (with every player on each team being in exactly the same physical condition as they had been initially, and with the information and experience acquired during the season all forgotten), then the same team each time would be the champion. Whether such a rollback were to occur one time or 100 times or 1,000 times or 10,000 times, a deterministic approach would always visualize the same winner. In effect, this is how a deterministic approach *must* visualize the operation of social processes. If the outcomes were not to be always the same under these assumptions, then any actual outcome would have to be treated as a member of a set of different outcomes compatible with the initial conditions rather than a uniquely inevitable product of those conditions.

From the probabilistic perspective adopted here, the actual championship team is not necessarily the best team. That is, if we likewise visualize the counterfactual condition of rolling time back to the beginning of the season in the way as described above, the same team would almost certainly not win in each try. To be sure, the distribution of winners would differ greatly from what would be expected under chance (which itself would produce a somewhat uneven distribution). But if the teams do differ in their potential to win the championship, some teams would almost never be winners, and others would win far more frequently than their share. Hopefully, the team that actually did win the championship is the team that most frequently wins in this series of rollbacks. Indeed, knowing only about the one actual series, we would have grounds to assume that the actual winner is the team most likely to be the most frequent winner in the rollback of 1,000 times. But this point needs to be put very carefully. It only means that the winner in a series of 1,000 rollbacks is the most likely winner if we have only one championship season. Depending on the distribution of teams, which is unknown to us, it could easily be the case that far in excess of 50 percent of the time the winning team is not the best team (as determined in a large series of rollbacks). Thus, what we mean by the "best team" is no longer what one typically means by such a phrase. Unless one team is extraordinary compared to any and all other teams—and this is going to be rare—we do not know which team in a league is the best as customarily defined. Since there are no rollbacks of 1,000 times, the best team is unknown, for the best team is the one with the greatest likelihood of being the winner in a series of such rollbacks. This likelihood might be almost unity or it could be far less than .5.

One could argue that this description of the deterministic perspective is too extreme, in requiring that the winning team would have to repeat in an infinite number of rollbacks. One might claim that this creates a straw man. Suppose, in a thousand trials, the same team wins, say, 995 times. Would that not be good enough? In point of fact, the deterministic perspective is incorrect if there is an expectation of anything less than an infinite stream of rollbacks with the same outcome. However, if the probability of the same team winning is virtually 1.0, then the deterministic viewpoint provides an outstanding approximation of the results obtained under a probabilistic approach. However, do we know this to be the case? On the other hand, do we know this *not* to be the case? We do know that the deterministic approximation of the probabilistic model will vary from situation to situation. Unless we have strong grounds for assuming a close approximation, we can make enormous errors by pursuing a deterministic approach.

A nice example of the role played by chance in such cases is the partially random interaction between the schedule of opponents and the health of the team. For example, an injury to a key player during the regular season might be of considerably less consequence if it occurs during a time when weak opponents are being played. If it occurs during a tough part of the schedule, the team's record for the season could be seriously affected. This is a good example because any variation in skills in avoiding injuries and any variation in skills in scheduling to optimize a team's outcome can only be loosely tied together, and hence chance factors will play a role in the combination of timing of injuries and opponents. But this is only one example; a somewhat more systematic set of observations is in order.

### *The Super Bowl*

Let us examine the issue in a somewhat more rigorous empirical manner. We will look at the winner of each of the twenty-seven Super Bowl professional football championships in the United States through 1993. In each case, we will start with the winner of each Super Bowl and ask how "certain" or "clean" was the team's progress to that championship. Now this is complicated. First of all, we are dealing with wall-to-wall counterfactuals of a heroic nature. If a winner barely edges out a loser because of the latter's misplay, for example, then we have no idea how the winner would have done if more effort had been required, i.e., if the misplay had not occurred. Secondly, it is reasonable to assume that



misplays are not random. In other words, some teams have better players or better training programs or better strategies that reduce their likelihood of misplays and/or increase the likelihood that the team will cause a misplay by its opponent. There are many examples: a superior defense may force the quarterback to throw inappropriately and thereby increase the chances of an interception; a well-executed tackle may be aimed at knocking the ball loose from the carrier thereby creating a fumble. Likewise, coaches may differ in their strategies for certain situations or their ability to detect a weak spot in an opposing player and thereby capitalize on it in a situation which the naïve observer views as simply a bad break. Certainly it is important to minimize the role of chance in influencing outcomes, and certainly it is not entirely separable from such matters. There is reason to believe that good teams make their own breaks. But it is also sound to recognize that to judge a team to be superior because it won a game by making its own "breaks" may well involve circular reasoning. Finally, there is a danger in a game with a close final score of finding the "chance" events which helped the winner, but ignoring those events which worked against them. For all we know a team may win by a narrow margin only because it was the victim of a string of bad breaks.

I shall focus exclusively on the Super Bowl champion teams. I will start with the Super Bowl game itself, asking if the championship game is won by a relatively narrow margin, seven points or less. If not, then we will ask the same question about the playoff game win which allowed the team to reach the Super Bowl (in other words, the semi-finals, which consist of the final two teams in the league). Again, if the margin is more than seven points, we go on to the previous round in the playoffs, and so forth. We then ask if the team got into the playoffs because its record was at least two wins better than the next team.[7] If none of these conditions occurred, we will conclude that the path to victory was not marked by any obvious chance events. Of course this set of criteria incorporates far less information than may genuinely be relevant. On the one hand, a more extensive analysis would also consider more carefully the factors affecting what teams end up playing the eventual winner. There is reason to think that chance factors will also affect which opponents are faced by the eventual Super Bowl winner—and therefore the actual opponents are not necessarily the strongest of the possible opponents. Likewise, we do not deal with how another team from the other division or league might have fared if they had won some playoff which could have led them to the Super Bowl against the winner. There are many events in the season that could be viewed (at least in part) as being accidental or chance occurrences, e.g., a key player (say the quarterback or the leading runner) is injured for the bulk of the season and therefore the team is knocked out of contention. On the other hand, as observed earlier, a close game (as defined here) might have been close only because the winning team had a lot of bad luck. Hence, the opposite conclusion would be justified. There are many circumstantial factors that are unanalyzed as well: for example, calls made by the game officials are sometimes arbitrary or difficult (for example, pass interference, or the exact point where the ball is placed after a tackle, or the exact point on the sideline where a punt crossed out of the playing field).

The analysis presented below is based on the data reported in David S. Neft, Richard M. Cohen, and Rick Korch (1993). I used their volume not only to trace the scores of all relevant games, but also as a source of descriptions of the Super Bowl and league championship games which they supply (usually, there is no text accompanying the results of the playoff games within the leagues). Keep in mind that the accounts of the games given in this source were presumably not written by the authors with the intent of underwriting my thesis here about a probabilistic approach to broad social events. Their accounts provide, therefore, an independent source of descriptive information about the games.

The following facts are not subjective: essentially half of the eventual winners won their Super Bowl game *and* their league championship *and* any preceding playoffs all by eight or more points, *and* initially entered the playoffs because they had a record two or more wins better than the team just below them (see chart 1). However, in thirteen of the twenty-seven seasons the margin of victory for the ultimate winner in the Super Bowl game was seven or less points in at least one of the post-regular season games: either in the Super Bowl game itself and/or in the league championship and/or in a playoff game leading to the league championship game. In five of these thirteen seasons, the margin of victory was close in more than one of these games. In Super Bowl XXV, for example, the New York Giants won the Super Bowl game by one point, 20–19, after winning their NFC Championship by two points, 15–13. In Super Bowl XVI, San Francisco won the Super Bowl by five points, 26–21, after winning its NFC Championship game by one point, 28–27. The most extreme example thus far is the Miami team that won Super Bowl VII by seven points, 14–7 after winning the AFC Championship 21–17, which in turn was preceded by a conference playoff victory of 20–14.

To be sure, a close score may still mean that the "better" team (as defined here in terms of the unknowable counterfactual outcomes of playing the season over and over again) was the winner. After all, a series of rollbacks in which one team was a fairly consistent winner would, from time-to-time, include some close games. In making such a point, however, keep in mind that the opposite would also hold. To wit, any given game might easily have been won by more than seven points (and hence would not be included here) even if the opponents were fairly evenly matched in a series of rollbacks.

**CHART 1.** Close Scores for the Team Winning the Super Bowl[a]

| *Super Bowl and Season* | *Playoff* | *Conference Championship* | *Super Bowl* | *No Post-Season Close Scores* |
|---|---|---|---|---|
| I 1966–67 | | 34–27 | | |
| II 1967–68 | | 21–17 | | |
| III 1968–69 | | 27–23 | | |
| IV 1969–70 | | 13–6 | | |
| V 1970–71 | | | 16–13 | |
| VI 1971–72 | | | | X |
| VII 1972–73 | 20–14 | 21–17 | 14–7 | |
| VIII 1973–74 | | | | X |
| IX 1974–75 | | | | X |
| X 1975–76 | | 16–10 | 21–17 | |
| XI 1976–77 | | | | X |
| XII 1977–78 | | | | X |
| XIII 1978–79 | | | 35–31 | |
| XIV 1979–80 | | | | X |
| XV 1980–81 | 14–12 | 34–27 | | |
| XVI 1981–82 | | 28–27 | 26–21 | |
| XVII 1982–83 | | | | X |
| XVIII 1983–84 | | | | X |
| XIX 1984–85 | | | | X |
| XX 1985–86 | | | | X |
| XXI 1986–87 | | | | X |
| XXII 1987–88 | 21–17 | 17–10 | | |
| XXIII 1988–89 | | | 20–16 | |
| XXIV 1989–90 | | | | X |
| XXV 1990–91 | | 15–13 | 20–19 | |
| XXVI 1991–92 | | | | X |
| XXVII 1992–93 | | | | X |
| TOTAL | 3 | 10 | 7 | 14 |

[a] Games won by seven or fewer points.

The descriptions of some of these games support the speculations suggested by the statistical pattern in the twenty-seven seasons thus far. For example, the Super Bowl V game, in which the Baltimore Colts defeated the Dallas Cowboys by 16–13 is described by Neft, Cohen, and Korch as "a comedy of errors" (1993, 207). Consider the close score and the wide range of misplays, such as fumbles and intercepted passes at key points, as well as critical injuries. Some examples (all from page 207):

> Baltimore quarterback Johnny Unitas threw a long pass down the center of the field to wide receiver Eddie Hinton; the ball bounced off Hinton's hands, back up into the air, grazed the fingertips of Dallas cornerback Mel Renfro, and came right down to the surprised John Mackey. Taking the ball around mid-field, Mackey sprinted the rest of the way to the end zone.

Or, consider the other side of the equation: a Colt player fumbled the opening kickoff in the second half

> deep in Baltimore territory. The Cowboys then drove from the 31-yard line to the two-yard line on five plays, with Thomas' hard running the key element. With the ball in the shadows of the goal posts, Thomas took a handoff and fumbled the ball, the Colts recovering on the one-foot line.

With the Cowboys ahead by seven points in the fourth quarter,

> a Morton pass [Morton being the Cowboy quarterback] bounced off the fingers of fullback Walt Garrison into the hands of Colt safety Rick Volk, who returned the ball 17 yards to the Dallas three-yard line. In short order, Tom Nowatzke smashed over for the touchdown, and Jim O'Brien added the tying extra point.

There are many other examples of such factors which, if not entirely random, are hardly to be viewed as non-accidental. Miami, the winner of Super Bowl VII by a 14–7 score, had first won its AFC Championship by 21–17 over Pittsburgh. Early in the first quarter of the AFC game, the Pittsburgh quarterback was knocked dizzy when he fumbled the ball and was out of the game until the final seven minutes (Neft, Cohen, and Korch 1993, 242).

Consider the following events in Super Bowl X, in which Pittsburgh defeated Dallas, 21–17:

> Through the first three quarters Dallas held a 10–7 lead. Then, at 3:32 of the final quarter, Reggie Harrison, a Pittsburgh reserve running back who plays on special teams, blocked a punt by Mitch Hoopes at the Dallas 9. The ball bounced off Harrison's face hard enough to wind up in the Dallas end zone, good enough for a two-point safety and run the score to 10–9. It was a play which was considered the turning point of the game. . . .
>
> Dallas coach Tom Landry blamed the defeat on the blocked punt by Harrison, which he said changed the momentum of the game around. He may have been right, but Swann's performance [Swann being a wide receiver for Pittsburgh]—which earned him the game's Most Valuable Player award—was momentum enough for the Steelers. Hospitalized only two weeks earlier with a concussion, and dropping passes in practice, the fleet-footed receiver returned to catch four passes for an astonishing total of 161 yards—a Super Bowl record certain to stand for many years. (Neft, Cohen, and Korch 1993, 297)



This last account is a helpful example of the operation of both skill and chance factors. A blocked punt represents a variety of skills, such that the blocker successfully evades those trying to prevent him from getting close enough to the punter, the inability of the punter to get the kick off quickly enough and/or at an initial arc which could not have been blocked. Presumably there was a combination of skills which worked to Pittsburgh's advantage, but the direction of a ball bouncing off the defender's face (hopefully the authors meant to say his face mask) is hard to allocate to a skill. Yet it did take a bounce such as to end up in the Dallas end zone.

Likewise, Swann, a pass receiver who was a key player in Pittsburgh's offense, is a victim of an initial bad break, to wit the concussion suffered two weeks earlier. Apparently, he recovered just in time to play a critical role in the offense. Had he recovered just a tad later (remember Swann was dropping passes in practice), the impact could have been great. Here again the events reflect a mixture of skill, the pass-catching ability of Swann, and of unfortunate chance events (the injury to Swann) coupled with a recovery which occurs just in time for the game (assuming recovery from a concussion is not simply a psychosomatic response to the urgency of the game). Injuries are inherent in the sport. Differences may occur between players in their proneness to injuries and their ability to play under less than optimal conditions. Likewise, differences may exist between teams in the exposure of their players to injuries. For example, teams will differ in the protection they are able to provide their quarterback and this, in turn, obviously affects the chances that such a key player will be tackled frequently and possibly injured.

I could provide many other examples of close games that could have gone either way and in which small "random" events seriously impacted the outcome.

The New York Giants defeated Buffalo by one point in Super Bowl XXV. In the final two minutes, Buffalo missed a 47-yard field goal attempt with a kick that went wide to the right (p. 627). Now a 47-yard field goal would be a real accomplishment, but it is not unheard of. We could view the matter as follows: what are the chances of the Buffalo kicker making such a field goal under the defensive conditions provided by the New York Giants? The actual failure could then be viewed in the context of these odds.[8]

Focusing on the issue at hand, in summary many of these close games suggest victories that might easily have been losses had not some events occurred which may be best described as having a random component. Contests are mixtures of luck and skill, no different than many card games, say, bridge or poker. Presumably random factors will wash out in the long run. But in small-N research there is no way to approximate this condition, and in large-N studies other factors obtrude to prevent us from ascertaining more than that certain events are probabilistically linked with others.

### *Social Relevance*

It might be thought that while random factors do play a role in sports outcomes, it is otherwise in the case of broad macrosocial events. Is it reasonable to also conceptualize large-scale social processes in terms of probabilistic forces? This is a less risky step than one might otherwise assume.[9] First, there is a non-trivial technical matter: with so many measurement errors in social research, a probabilistic model is in order even if one wants to be a determinist. Second, there is evidence that a probabilistic approach is the best that can be achieved in many areas of the "hard sciences," where the influences on some outcome are best described in terms far weaker than would be required on a deterministic model. See, for example, the discussion in Lieberson (1985, 226–27) of the "amplification effect" described by Weisskopf as operating in a variety of scientific contexts where very minor causal perturbations sometimes have very large effects; or the review of probabilistic issues by Lieberson (1992, particularly pp. 7–9), the discussion by Salmon (1971, 56–57) of "incomplete explanations" of paresis and radioactive decay, as well as the literature cited in these publications. Finally, there are grounds for taking a probabilistic perspective even if one is convinced that a deterministic model is actually operating, and that everything allocated to the residual category of "chance" consists of nothing more than unexplained deterministic process.

The rationale for this latter claim rests on a simple fact: in social as well as in many other complex processes, elucidating all of the factors that actually contributed to a given outcome would be highly counterproductive given the goal of generalizable knowledge. Even if it were possible to do so, the massive effort



required would involve assessing the roles of all sorts of factors whose occurrence or failure to occur in a particular instance would be wholly incidental from the viewpoint of seeking to understand a particular type of process. Suppose we take a deterministic view of history such that a given outcome, say a war or the assassination of a prominent political figure, *had* to happen when and how it did. What knowledge would it take to locate and understand every conceivable influence for and against that outcome's occurring in the exact time and place that it did? It would be, for all practical purposes, infinite. We would need to account for everything from the birth of the assassin, the life history leading to a desire to kill the victim, to the decision of the political figure to go to the specific place where he or she was assassinated, to the availability of the weapon used, to the behavior of the Secret Service agents, and so forth. Surely it would be more reasonable to renounce such a goal, recognizing that complex chains of events in their particularity cannot be subsumed under generalizable laws or theory (more on this below). In examining a single case, or when engaged in a small-N study, how confident can we really be that the observed outcome(s) would be repeated over and over again?

The role of random factors is, of course, relevant to a wide variety of social contexts. Consider a simple example: undergraduate admission to Notre Dame or Harvard, or to any other selective institution with more "qualified" applicants than it can accept. If a highly selective college were to reconsider the same set of applicants for admission, using the same admissions personnel, and following the same guidelines, there would presumably be a correlation between the admissions at time 1 and time 2. But it would hardly be a correlation of 1.0 . Some who were admitted in the first review would be denied admission on the second trial; in other cases, the opposite would occur. There would be a huge population denied admission on both trials (a numerical necessity when the number of admissions is much smaller than the number of applicants) and there would be many who would be admitted in both cases because of exceptionally distinguishing characteristics. But the key point is that there would be a certain number of winners who fall out in time 2, and a certain number making it at time 2 who were denied admission at time 1. This is interesting to me for two reasons. First of all, we can see that there is a certain random dimension to the process that is being ignored when we use admission (like, victory) to settle questions about who are the "best applicants" (or "best teams"). Most interesting are the questions that inevitably arise here about the specific features of the structures involved. That is, there is no reason to assume that correlations among trial outcomes would converge on a specific value in all similar systems of this sort. Systems could vary because of the nature of the judges, the nature of the criteria, the distribution of qualities among the applicants, and the ratio between the number to be admitted and the number of applicants. Also, except for sampling errors, we would expect the characteristics of the aggregate of successful applicants to remain approximately constant from trial to trial, even if the same applicants were not always successful. So there is much that could be done here to explore the relative role of chance factors and the ways in which the features of particular structures influence this role.

There is a second direct application. When there is an elimination tournament such that a team is out when it loses one game (as is the case for the Super Bowl playoffs), then we have an irreversible dependent variable (winning or losing). This model has relevance insofar as there are social events that we want to understand which have the same quality, e.g., assassinations or wars or the emergence of social movements. From the point of view of social processes, there is every reason to think that the analogy holds. Just as a team's outcome is a function of skills as well as of chance events, so too social processes involve both complex and recurring structures and highly contingent events—and the parallel is particularly significant when outcomes are not directly reversible, in the sense that the absence or removal of whatever factors are responsible for a particular outcome will not (or not alone) enable the situation to return to its previous state (see the discussion of asymmetrical forms of causation in Lieberson 1985, chapter 4).

An intercepted pass or a fumble is the product of an interaction between the skills of one team's offense, the skills of the other team's defense, and chance events. Recurring social structures operate as do such skills to increase or decrease the probability of certain outcomes. But as in the sports event there are always other factors, best called chance factors, which may impact on social outcomes. The limits within which such factors may operate to affect decisively a particular outcome will vary, as will the importance of what we have characterized as "chance." But this variability, too, is mirrored in the differing roles played by chance *vs* skill in different card games and in different sports (depending on both the sport *and* the rules). The ramifications of these analogies for understanding social processes are enormous, as I will attempt to show below.

## Some Proposals

On the basis of our analysis thus far, the remainder of this essay sketches out some proposals for how to think about major social processes. We have started by reviewing some grounds for dissatisfaction with current methodologies used on small-N data sets as well as with the deterministic mode of thinking which underlies them. In a certain sense the description of small-N research also revealed difficulties for large-N research and conceptualization as well. It is just that the difficulties of the latter are compounded when causal inferences are drawn on the basis of small N's.

### *A Multi-layered Probabilistic Approach to Social Processes*

I start with three assumptions. First, there is no alternative to a probabilistic perspective—even if we believe social causes operate in a deterministic mode. The second assumption is that there will always be too many different causal influences for us to take all of them into account at one and the same time. My third assumption is that not all causal influences operate in the same way. By that I mean that it will frequently be useful to think of causal factors as hierarchically organized such that some set the precondition for others to operate and that the latter, in turn, regulate what further factors come into play. Ultimately, of course, we must reach factors capable of directly influencing the behavior of individuals. We might visualize "layers" of such causal principles (i.e., formulations of relatively stable relationships among causal variables) in the following way. Suppose we have a given principle, $P_2$, and there are certain conditions which determine its operation or non-operation. Then principles which determine these conditions (say, $P_4$, $P_7$, $P_{14}$) are at a higher level than $P_2$. On the other hand, if $P_2$ represents a relationship affecting conditions which in turn influence the operation of some other principle, say $P_1$, then $P_2$ is, in that respect, at a higher level than $P_1$. A fourth assumption is that genuine stability in social processes is rare; in modern societies at least, the conditions affecting the operation of social causes are forever subject to change along with their influence on lower level processes and events.

If these assumptions are correct, then in seeking to establish principles expressing causal influences we shall have to recognize the following:

1) These principles will typically be expressible only in probabilistic terms.
2) There are bounds on the operation of each principle, and that one task will be to specify the conditions under which particular principles operate. Indeed, a principle may even be totally irrelevant in a specific setting at a specific time. The principle is not true or false, per se, if it fails to hold in a given setting. It is false only if it *generally* fails to hold in a setting in which we have reason to believe it should operate (more on this below in discussing Central Place Theory) or, conversely, operates when its bounds predict it should not.
3) If a principle is to be useful, we must be able to state the bounds or conditions under which it will or will not operate. Universal social principles, intended to apply without limits of time and place, are extremely

unlikely (other than those which are biologically determined). Hence, no principle is acceptable unless it is possible to provide a rough approximation of the assumptions underlying its operation. It is impossible for such bounds to be fully tested and they may even prove to be impermanent, but one starts with such limitations and then adjusts as the future unfolds and the principle's application to the present and past is expanded (into, say, different places, settings, combinations of conditions, and so forth).

The various permutations and combinations of principles which can tug against each other in conflicting directions or pull together in tandem towards a given outcome makes for an interesting set of problems, namely figuring out the conditions under which the principles operate both individually and jointly. Under which conditions will one triumph over the other, or are the outcomes simply intermediate between the two?

Such questions, however fascinating, cannot be further explored here. Rather, I will briefly describe several cases that illustrate what I intend in calling for a multi-layered probabilistic approach to the conceptualization of social phenomena.

1) There is good reason for believing that race riots have both underlying causal conditions and immediate precipitants. The immediate precipitants are events which trigger a riot. Examples are: abusive interaction between police and members of a given racial or ethnic group; members of one group who assault someone in another group (particularly, if the latter is a small child, woman, elderly, or otherwise someone for whom there are rules of special care and concern, such as someone who is blind, highly regarded in the community, a member of the clergy, a person risking his or her life for the good of the community). The nature of the action is also relevant for the immediate precipitant, for example, a rape is more provocative than a purse snatching. At any rate, these events occur all the time, but do not always lead to riots. Indeed the vast majority of such incidents do not. When do riots occur? There are certain institutional conditions which appear to affect the likelihood that such painfully common incidents will increase or decrease the chances of a riot (see Lieberson and Silverman 1965). For the purposes at hand, we visualize two principles operating. One is a principle dealing with events which trigger riots, the other—at a higher level—deals with conditions which increase or decrease the chances that such incidents will trigger a riot. We can visualize two separate switches—if both are turned on, the chances of a riot are greater than if either is turned off. If both are turned off, then the chances are extremely low (observe, incidentally, that I am describing this in probabilistic form to begin with—not in flat terms such as: there will or there will not be a riot). If the precipitating incident switch is on, but the underlying condition switch is not, then the chances of a riot are extremely modest; likewise, if the opposite conditions hold.



2) Another race and ethnic example from my favorite author. In a theory of race and ethnic relations that I developed many years ago, I drew a distinction between two types of ethnic/racial subordination: groups that migrated into such situations from another setting *vs* groups that were conquered or otherwise overrun and thereby became subordinated. I argued that different consequences flowed from these conditions. Of interest here is the claim that separatism and reduced levels of assimilation were far more likely to occur when groups were conquered than when groups entered subordination through migration.[10] I think this proposition holds—separatist movements are far more likely to occur among conquered peoples, and these people are likely to be slower to assimilate. But not all conquered groups in such subordinated settings develop separatist movements. So there are several interesting possibilities: 1) that is the best we can do; we cannot *a priori* separate the two subsets of conquered subordinate groups into those who do or do not push towards separatism; 2) in addition, for those who do push towards separatism, can we offer any principles as to when they will do so? Although the essay does not answer this, it is easy to see that a number of relevant lower levels of principles could be operative. To wit, given the conditions of inter-ethnic contact, what will generate the next stage and when? Likewise, we have the need for higher-level principles as well: to wit, under what conditions will groups migrate into subordination? Under what conditions, will one group conquer another?[11]

3) The third example, based on Central Place Theory as propounded by Lösch (1954), I have discussed in greater detail elsewhere (Lieberson 1992, 8). Basically this is a theory about the distribution of cities of different sizes and their distances from one another. It fails to go very far in helping us understand the spatial distribution of different-sized cities in the United States. The location of New York, for example, is off by about 1,000 to 1,500 miles! It is, by my tastes, an elegant and sensible theory and works beautifully in some specific instances. The key fact is that Lösch has carefully stated the conditions under which his model operates. When these conditions are met, the model seems to work very nicely. When these conditions are not met, then the principles he proposed carry very little weight (although I suspect that over time we would still see movements in the direction of his model). The point is that the Central Place Theory formulated by Lösch includes its own preconditions. We thus are given the higher-order principles that set the limitations on when this theory will work or not work. This is, then, an excellent illustration of the explicit use of multi-layered principles in social theorizing.

4) The first names given to children—a topic I am currently studying—provides another example of the potential fruitfulness of an approach to theorizing which emphasizes the multi-layering of principles. At this point I speculate that certain broad social conditions will impact on the likelihood that different names will be given to children. And, in turn, this makes it possible for lower-level principles of taste to enter into the setting. For example, the feminist movement (which itself is a product of the operation of other principles) influences the kind of names that are acceptable or unacceptable to parents when they name daughters. If certain names decline in popularity because of the feminist movement, then new names will replace them. What new names will they be? Certainly, the feminist movement will influence the choices, but there is now room for other principles of taste to affect the new choices that are made from within the boundaries or limitations set by the feminist movement. Hence the specific choices will reflect a lower-level principle that is activated by the condition caused by the higher-level principle.

5) Finally, let us think through the meaning of the phrases "an accident ready to happen" or "the time is ripe" or "not ripe." Each of these phrases suggests that there is some set of conditions which is necessary for something else to happen: say, a war, or a social movement, or a major institutional change, a political shift, etc. The implication is that some such set of conditions plays a critical role in either impeding or making possible a particular outcome. The outcome has a certain independence in that the "higher-level" conditions do not guarantee its occurrence or non-occurrence, but clearly a point can be reached where lower-level events are almost certain to trigger the outcome simply because the relevant "triggers" are extremely common, and/or because the probability is extremely high that the outcome will occur whenever potential triggering events are encountered. So we can develop a wide variety of conditional linkages between the sets of layers.

## Chain Models

If important aspects of social life are accurately describable in terms of multi-layered processes of the sorts discussed above, and if there is multicausality for just about any given dependent variable, then it seems reasonable to assume that many of our questions will require the unraveling of complex probabilistically linked chains of events rather than simply searching for "the" cause or causes of some phenomenon of interest. This means that many questions about society, which may initially sound quite sensible, in reality have very limited meaning. If we try to identify a causal connection between two broad and significant characteristics of society, say the influence of some "key" economic condition on certain major political characteristics, or perhaps the effect of a social characteristic (say linguistic or ethnic diversity) on some economic characteristic, there is a good chance that a variety of causal pathways exists between two such phenomena, each one conditioned by different predisposing circumstances. Moreover, such pathways themselves may lead to different outcomes when activated. Hence some pathways from, say, key economic conditions to a major political characteristic may move the latter in one direction, other paths in the opposite direction, and yet others serve to maintain the political status quo.



Now obviously we can choose as our independent variable one "so close" to the dependent variable that there will not be much of a chain. But few interesting questions could be answered by adopting that strategy. For example, suppose I want to account for the social characteristics of those who become physicians. Where do I start? If I start very late in the process, say, with the behavioral orientations of those who are already in their second year of medical school, I am unlikely to learn anything of much interest. If I start with people just finishing high school—or even earlier—I am going to find a very complex chain of events leading to differentials in who becomes a physician. There are going to be a lot of points in this complex set of paths to consider, though some, to be sure, will prove to be more important than others.

Many links in the chains may be very weak—weak in the sense that a movement from one point to a distant point may occur through a variety of connections each of which has a low probability of occurring. In other words, it may be that nothing can be said with certainty about what outcomes will occur—particularly if we do more than go back to the equivalent of second-year medical students. This is the sociological equivalent, as it were, of the Super Bowl process. There are a lot of "chance" influences at work. If we start by asking who wins the Super Bowl game, and take as given that the two actual finalists are playing each other, then we are dealing with something closer to the case of second-year medical students. If we go back to the beginning of the season, that is, if we do not take for granted who the finalists are, then we have a much more complex and less certain chain of events. It has to be more likely that the best of the two finalists will win the Super Bowl than that the best team at the start of the season will win it. Indeed, this is definitional, since the chances of the latter can be no higher than the chances of the former (and easily considerably lower).[12]

This view of long chains fits nicely with the position taken by Popper (1964, 115):

> although we may assume that any actual succession of phenomena proceeds according to the laws of nature, it is important to realize that

> practically *no sequence of, say, three or more causally connected concrete events proceeds according to any single law of nature.* If the wind shakes a tree and Newton's apple falls to the ground, nobody will deny that these events can be described in terms of causal laws. But there is no single law, such as that of gravity, nor even a single definite set of laws, to describe the actual or concrete succession of causally connected events; apart from gravity, we should have to consider the laws explaining wind pressure; the jerking movements of the branch; the tension in the apple's stalk; the bruise suffered by the apple on impact; all of which is succeeded by chemical processes resulting from the bruise, etc. The idea that any concrete sequence or succession of events (apart from such examples as the movement of a pendulum or a solar system) can be described or explained by any one law, or by any one definite set of laws, is simply mistaken. There are neither laws of succession, nor laws of evolution.

Even if fairly high probabilities operate to link points along an actually observed chain of events, the likelihood of any chain of events repeating itself with any reasonable frequency is minuscule when the chain is long. Under those circumstances a theory concerning a complex set of events predicated on very specific outcomes is not possible. Or, if developed for a single case, has no applicability to the general problem. If we could replay history, which is no more possible than replaying a football season, then the probabilistic view is that we would encounter a rather different history each time. Presumably some outcomes would occur more often than other outcomes, but the distribution of outcomes also would depend on what specific types of events we had chosen to study. Consider the different expectations we might have if we started with, say, manners in 1800 in the United States and then looked at manners in 1900, or slavery in 1800 *vs* 1900, or a country's form of government over an extended period, and so forth. The outcome distribution also might depend on how precisely and concretely we identify an outcome of interest: a specific manner *vs* a theme in manners (see, for example, Elias 1978); a specific feature of government as opposed to the general degree of democracy, etc.

To be sure, one can visualize points on some chains where the path to a given outcome is not preordained, but still where a set of interrelated outcomes all point to a certain final event. Consider the comments in the preceding section about an accident waiting to happen, the time being ripe, etc. Hence, the system of chains is an interesting problem to consider in understanding social events. As is the case with some diseases, in which the body is so weakened that one or another, otherwise minor, illness is likely to prove fatal, so too, we can at least visualize social processes proceeding in such a way that at some point there is a very strong likelihood of a certain outcome, even if the particular path to that outcome has a probability far less than 1.0. Of course, this will always be an empirical question rather than an assumption we can ordinarily make. I think it unlikely, for example, that long complex chains of causal influence in the social realm will have such "absorbing" qualities, except perhaps close to the end of a relevant sequence (which of course is what Popper is in effect claiming in the quotation above).



To diverge briefly from the main point, this analysis also implies something about second guessing. The problem is not that the second-guesser uses 20/20 hindsight to reach the correct solution—rather it is that the original plan may well have been the absolutely best one even though it failed. This is because such efforts are often based on complex probabilistic chains of reasoning. Hence the best shot is simply the one which seeks to activate the most reliable of what could be a set of low-probability chains of causal influence.[13] Incidentally, this illustrates a difference between the probabilistic and deterministic approaches to policy issues. The latter says there is a correct way, which when followed, will produce the desired results. If the desired result does not ensue, then some error was made. A probabilistic view may acknowledge that one action is more likely to produce the desired outcome than any others, while yet recognizing that the outcome is not inevitable; it simply has a higher probability. The failure of a policy to produce a desired outcome thus does not imply that it would be inappropriate to seek to implement it again in the future, so long as it was based on a well-confirmed theory and adequate data. However, such an approach may be politically very difficult unless both decision makers and those likely to be most affected by the decision are prepared to take a probabilistic view of failure.

## Explaining Specific Events

As a general rule, there are limitations to how well we can explain (or understand the cause[s] of) a specific social event. These limitations exist even if the research is based on very large N's. There is the difficulty of applying existing theory (regardless of whether it is derived from small- or large-N research) to help us understand the cause of a specific event. In effect, the explanation of a specific event is a special small-N problem even if the knowledge applied to it draws on large-N research. When we deal with a specific event and its explanation, the total number of all such events is immaterial since we are only concerned with one of them. In small-N research we are always dealing with a small number of events; but if we are trying to explain one specific instance of a type of event, then there are special difficulties regardless of whether there are many such events or only a few—we are still trying to account for only one specific event.

At best, we can come up with a probabilistic statement of the causes of a given event—a probabilistic statement which normally will involve two parameters, and very likely more than two. It will almost inevitably involve a stochastic variable as well as at least one known causal variable because, typically, research will show that the substantive variable fails to fully account for the dependent variable. More often we will have good reason to believe that additional substantive variables are operating (the multivariate case) if for no other reason than our usual inability to experiment or otherwise *situationally* measure or control other influences.

Visualize the simplest case in which there is one independent variable and a stochastic variable ($X_1$ and $U$, respectively), and where both the independent variable and the dependent variable are dichotomous (simple yes/no variables). If there is a relationship such that "yes" on the independent variable is associated with the occurrence of the dependent variable, can we say that the presence of $X_1$ is the cause of Y? No, we cannot, even though this is a common form of thought. Suppose we know that there is a causal linkage between smoking and lung cancer. If someone we know dies of lung cancer and was a heavy smoker, it is hard for us to avoid thinking we know the cause of this particular individual's lung cancer. But if there are non-smokers who die of lung cancer, then we really cannot explain the cause of the specific event—the death from lung cancer of our smoker-friend. Depending on the proportion of all smokers who die of lung cancer in comparison with the proportion of all non-smokers similarly afflicted, I can make some statements about the relative probability of the death's being due to smoking—but no more unless there is additional information. In other words, the excess in the rate of lung cancer deaths for smokers compared to non-smokers presumably tells us about the increased likelihood of fatal cases of lung cancer due to smoking. But it does not tell us about the cause of any particular smoker's lung cancer. In counterfactual conditional terms, unless we know a lot more about other causes of lung cancer death, we can safely say that some percentage of smokers' deaths would not have occurred if they had not smoked, but we cannot say anything about any specific smoker who dies of lung cancer other than to list the probabilities of different causes. Although I can do this very nicely for a population, I cannot really explain any individual case.[14] Notice how different this is from the very clean and clear statement about the probability of lung cancer for those who smoke *vs* the probability for non-smokers.

Suppose persons residing in a given area were subjected to exceptionally large doses of radioactivity (say due to chemical wastes or nuclear testing, etc.). Suppose those now dying from a cancer known to be caused by radioactivity



decide to sue those responsible for the radioactivity. Again all we can do is say that a certain percentage of these victims (possibly an extremely large percentage) are dying for this reason. But unless the probability of getting this cancer is otherwise zero, we cannot say why any given person is dying. We can only make a probability statement about the likely cause. The key is not whether all people subjected to radioactivity get the cancer, but rather whether there are people not subjected to this level of radioactivity who get that type of cancer. If only radioactivity can cause cancer, we have to that degree an explanation.[15]

I do not believe these examples concerning illness and disease are at all difficult to appreciate. But I have a hunch that we are likely to initially draw a conclusion unsupported by present knowledge whenever we hear of a smoker who died of lung cancer. The temptation to draw an unwarranted conclusion is even more difficult to resist when we have no way to evaluate the likelihood of alternative outcomes, or information is unavailable concerning the likelihood of the same outcome under different circumstances.[16] But the examples so far are less obtuse than examples involving large-scale social events: wars, macro-level societal changes, institutional developments, forms of social organization, political events, and the like. In the typical research problem in nonexperimental social science one will virtually always have reason to believe that more than one independent variable is operating. Here, if there is also a stochastic factor operating, only a set of probability statements can be made about the cause(s) of a specific event. Thus only in an incredibly unlikely set of circumstances would it be possible to talk about *the unique, necessary and sufficient cause* of a specific event. Suppose we complicate matters by assuming that there are two known causes ($X_1$, $X_2$), both being continuous variables, that the dependent variable, $Y$, is also a continuous variable, and that the relationship between each variable is linear. Then for any individual case, knowing the values of $X_1$ and $X_2$ and $Y$, we could only estimate the likely causal influence of each relevant variable. Here it is even easier to fall into the mistake of thinking that we can find the definite cause of each specific outcome. Unlike the medical examples used above, it is harder to visualize these sorts of cases as being probabilistic in nature and hence to recognize that it may be extremely difficult, and in many cases impossible, to say what really caused a specific event to occur.

## In Conclusion

Returning to the main point, in brief we should be suspicious of pied pipers who purport to tell us why such and such *had* to happen. This applies "in spades" to customary small-N analyses which seek to simulate the sort of results which might be attainable with large-scale data sets.[17] But it also applies to large-scale data analysis as well. The conjectures I have made here about how social processes operate may or may not prove to be wholly adequate; they are certainly open to correction and revision. However, I believe they provide a viewpoint from which it may be possible to pursue our questions in more fruitful and realistic ways, given both the obvious complexity of social phenomena and the limitations of our methods. Above all we should resist seeking a precision that is inappropriate for either our subject matter or the nonexperimental tools with which we must work.

## NOTES

I am indebted to the following for helpful discussions of some of the problems raised in this paper: William Alonso, John Campbell, Zvi Griliches, S. M. Miller, and Debra Minkoff. Angelina C. Wong assisted in the analysis of Super Bowl data.

1. For some problems, 100 cases is more than sufficient; for others a data set ten times that number will be inadequate.
2. Ragin and Becker (1992) provide a general discussion of the various ways in which a case can be defined in social research, and the consequences of doing so.
3. This is oversimplified and hardly does justice to the theory, but it serves to illustrate the way in which variance is forced.
4. Sometimes practitioners of this approach pool two variables together such that a certain level with one variable and a specific level with the other are viewed as combining to cause an outcome, whereas neither alone is a cause. But this is not really two causes, in the sense that either of two different conditions will lead to the same outcome. Rather it is one cause involving two separate conditions.
5. The Boolean solution proposed by Ragin (1987), although appealing in its ability to lay out all of the permutations and combinations of variables in relatively small-N situations, is still ultimately a deterministic approach.
6. The comparison between deterministic and stochastic models used in statistical research is relevant here. See, for example, Hanushek and Jackson (1977, 11–13).
7. As a general rule, it was determined if the eventual Super Bowl winner got into the initial playoff rounds with only a one game edge over the next best team.
8. To be sure there are unmeasured intangibles such as the kicker choking up or, at the other extreme, having a clutch player who does better with such incentives.
9. Keep in mind that the deterministic model is also an unproven assumption.
10. In fairness, the original paper (Lieberson 1961) did not describe these principles in probabilistic form, but I would now restate them in this fashion.
11. Some of these higher-principle questions are at least partially answered in Lieberson (1961).
12. The Super Bowl analysis of chance factors, reported earlier, is therefore relatively conservative since it did not consider the far more complex model.



13. If they are all more or less similarly low in probability, then substantial differences in the probable harmful effects of failure may be a more important basis for deciding than the small differences in the chances of success (a cost-benefit issue).
14. All of this assumes that the usual factors of age, sex, frequency of smoking, occupation, and the like are taken into account.
15. Incidentally, if not all people who smoke die of lung cancer and if not all subjected to radioactivity get leukemia, then there is the added question of why some do and some don't. This is inevitably intriguing, but it is not always answerable and may in fact reflect factors best allocated to randomness as well as to more definite influences (again see Salmon 1971, 56–57).
16. To use a sports analogy again, if the manager orders a given action which backfires, say a runner at a key point in a baseball game is caught attempting to steal a base, or if a football team unsuccessfully goes for a first down rather than punting, the usual assumption is that an error was made. Yet in such cases we do not have anything like the equivalent of data on lung cancer among non-smokers. Hence, we are in no position to say that the failure shows it was a bad call. We have no real grip on relevant counterfactuals, and the relevant sets of probabilities are clearly unknown.
17. I exclude here research which really can build on single instances through the use of exceptionally strong quasi-experimental conditions, such as the work by Snow on cholera (described in Freedman 1991).