Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals

By Jasjeet S. Sekhon

In contrast to statistical methods, a number of case study methods—collectively referred to as Mill's methods, used by generations of social science researchers—only consider deterministic relationships. They do so to their detriment because heeding the basic lessons of statistical inference can prevent serious inferential errors. Of particular importance is the use of conditional probabilities to compare relevant counterfactuals. A prominent example of work using Mill's methods is Theda Skocpol's *States and Social Revolutions.* Barbara Geddes's widely assigned critique of Skocpol's claim of a causal relationship between foreign threat and social revolution is valid if this relationship is considered to be deterministic. If, however, we interpret Skocpol's hypothesized causal relationship to be probabilistic, Geddes's data support Skocpol's hypothesis. But Skocpol, unlike Geddes, failed to provide the data necessary to compare conditional probabilities. Also problematic for Skocpol is the fact that when one makes causal inferences, conditional probabilities are of interest only insofar as they provide information about relevant counterfactuals.

> "Nothing can be more ludicrous than the sort of parodies on experimental reasoning which one is accustomed to meet with, not in popular discussion only, but in grave treatises, when the affairs of nations are the theme. . . . 'How can such or such causes have contributed to the prosperity of one country, when another has prospered without them?' Whoever makes use of an argument of this kind, not intending to deceive, should be sent back to learn the elements of some one of the more easy physical sciences."
>   —John Stuart Mill[1]

Case studies have their own role in the progress of political science. They permit discovery of causal mechanisms and new phenomena, and can help draw attention to unexpected results. They should *complement* statistics. Unfortunately, however, case study research methods often assume deterministic relationships among the variables of interest; and failure to heed the lessons of statistical inference often leads to serious inferential errors, some of which are easy to avoid.

The canonical example of deterministic research methods is the set of rules (or what are often called canons) of inductive inference formalized by John Stuart Mill in his book *A System of Logic.*[2] Mill's methods have greatly influenced generations of researchers in the social sciences.[3] For example, the "most similar" and the "most different" research designs, often used in comparative politics, are variants of Mill's methods.[4] But these methods do not lead to valid inductive inferences unless a number of very special assumptions hold. Some researchers

seem to be either unaware or unconvinced of these methodological difficulties, even though Mill himself clearly described many of their limitations.

Mill's and related methods are valid only when the hypothesized relationship between the cause and effect of interest is *unique* and *deterministic.* These two conditions imply the absence of measurement error, because in the presence of such error, the relationship would cease to be deterministic. These conditions strongly restrict the methods' applicability. When Mill's methods of inductive inference are not applicable, conditional probabilities[5] should be used to compare the relevant counterfactuals.[6]

The importance of comparing the conditional probabilities of relevant counterfactuals is sometimes overlooked by even good political methodologists. Barbara Geddes, in an insightful and often assigned article on case selection problems in comparative politics, neglects this issue when discussing Theda Skocpol's book *States and Social Revolutions.*[7] Skocpol explores the causes of social revolutions, examining the ones that occurred in France, Russia, and China, as well as the fact that revolutions did not occur in England, Prussia/Germany, and Japan. Geddes seriously questions Skocpol's claim of a causal relationship between foreign threat and social revolution.[8] Geddes's evidence is compelling if the only possible relationship between foreign threat and social revolution is deterministic—i.e., if foreign threat is a necessary or sufficient cause of social revolution. But she never considers the possibility that the relationship may be probabilistic—in other words, that foreign

Jasjeet S. Sekhon is associate professor of government at Harvard University (jasjeet_sekhon@harvard.edu). He thanks Walter R. Mebane, Jr., Henry Brady, Bear Braumoeller, Shigeo Hirano, Gary King, John Londregan, Bruce Rusk, Theda Skocpol, Suzanne M. Smith, Jonathan N. Wand, the editors of Perspectives, *and three anonymous reviewers for valuable comments and advice.*

threat may increase the probability of social revolution but not necessarily cause a revolution. This omission is of some importance because her data actually support Skocpol's hypothesized relationship if it is interpreted to be probabilistic.

My discussion does not, however, in any way undermine Geddes's criticism of Skocpol's research design for selecting on the dependent variable. In fact, Skocpol failed to provide the data necessary to compare the conditional probabilities.

Skocpol clearly believes she is relying on Mill's methods. She states that "[c]omparative historical analysis has a long and distinguished pedigree in social science" and that "[i]ts logic was explicitly laid out by John Stuart Mill in his *A System of Logic*."[9] Further, she asserts that she is using *both* the Method of Agreement and the more powerful Method of Difference.[10] For these methods to lead to valid inferences, though, there must be only one possible cause of the effect of interest, the relationship between cause and effect must be deterministic, and there must be no measurement error. If these assumptions are to be relaxed, random factors must be accounted for. And because of these random factors, statistical and probabilistic methods of inference are necessary.

The key probabilistic idea upon which statistical causal inference relies is conditional probability.[11] But conditional probabilities are rarely of direct interest. When making causal inferences, we use conditional probabilities to learn about counterfactuals—e.g., would social revolution have been less likely in Russia if Russia had not faced the foreign pressures it did? We have to be careful to establish the relationship between the counterfactuals of interest and the conditional probabilities we have managed to estimate. Researchers too often fail to do this and assume that the conditional probabilities they have estimated are of direct interest.

In this article, I outline Mill's methods, showing the serious limitations of his canons and the need to formally compare conditional probabilities in all but the most limited of situations. I then discuss Geddes's critique of Skocpol and posit several elaborations and corrections. I go on to show how difficult it is to establish a relationship between the counterfactuals of interest and the estimated conditional probabilities. I conclude that case study researchers should use the logic of statistical inference and that quantitative scholars should be more careful in how they interpret the conditional probabilities they estimate.

Mill's Methods

The application of the five methods Mill discusses has a long history in the social sciences. I am hardly the first to criticize the use of these methods in all but very special circumstances. For example, W. S. Robinson, who is well known in political science for his work on the ecological inference problem,[12] criticized the use of Mill-type methods of analytic induction in the social sciences.[13] Robinson did not, however, focus on conditional probabilities, nor did he observe that Mill himself railed against the exact use to which his methods have been put.

Adam Przeworski and Henry Teune advocate the use of what they call the "most similar" design and the "most different" design.[14] These designs are variations on Mill's methods. The first is a version of Mill's Method of Agreement, and the second is a *weak* version of Mill's Direct Method of Difference. Although Przeworski and Teune's volume is more than 30 years old, their argument continues to be influential. Indeed, a recent review of qualitative methods makes direct supportive references to both Mill's methods and Przeworski and Teune's formulations.[15]

Mill describes his views on scientific investigations in *A System of Logic*, first published in 1843.[16] In an often cited chapter (book III, chapter 8), Mill formulates five guiding methods of induction: the Method of Agreement, the Direct Method of Difference, the Double Method of Agreement and Difference, the Method of Residues, and the Method of Concomitant Variations. Some consider the Double Method of Agreement and Difference to be merely a derivative of the first two methods. This view is not accurate, because—as I explain later in this article—there is actually a tremendous difference between the combined method (what Mill also calls the Indirect Method of Difference) and the Direct Method of Difference. Both the Method of Agreement and the Indirect Method of Difference are limited and require the machinery of probability in order to take chance into account when considering cases where the number of possible causes may be greater than one.[17] Factors not well explored by Mill, such as measurement error, also invalidate these methods.[18] Even in the case of the Direct Method of Difference, which is almost entirely limited to the experimental setting, chance must be taken into account when measurement error is present or when interactions among causes lead to probabilistic relationships between a cause, A, and its effect, a.

Here, I will review Mill's first three canons and show the importance of taking chance into account as well as comparing conditional probabilities when chance variations cannot be ignored.[19]

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 agree is the cause (or effect) of the given phenomenon."   —John Stuart Mill[20]

A possible cause—i.e., antecedent—may consist of more than one event or condition.[21] For example, permanganate ion with oxalic acid forms carbon dioxide (and manganous ion). Separately, neither permanganate ion nor oxalic acid will produce carbon dioxide; but if combined, they will. In this example, the antecedent, A, may be defined as the presence of both permanganate ion and oxalic acid.

Let us assume that the antecedents under consideration are A, B, C, D, E, and the effect we are interested in is a. Suppose that in one observation we note the antecedents A, B, C, and in another we note the antecedents A, D, E. If we observe the effect a in both cases, we may conclude, following Mill's Method of Agreement, that A is the cause of a. We conclude this because A is the only antecedent that occurs in both cases—i.e., the observations agree on the presence of A. When using this method, we

seek observations that agree on the effect, *a*, and the supposed cause, *A*, but differ in the presence of other antecedents.

Direct Method of Difference

> "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon."
> —John Stuart Mill[22]

In the Direct Method of Difference, we require, contrary to the Method of Agreement, observations that are alike in every way except one: they differ in the presence or absence of the antecedent we conjecture to be the true cause of *a*. If we seek to discover the effects of antecedent *A*, we must introduce *A* into some set of circumstances we consider relevant, such as *B*, *C*; and having noted the effects produced, we must compare them with the effects of *B*, *C* when *A* is absent. If the effect of *A*, *B*, *C* is *a*, *b*, *c*, and the effect of *B*, *C* is *b*, *c*, it is evident, under this argument, that the cause of *a* is *A*.

Both the Method of Agreement and the Direct Method of Difference are based on a process of elimination. This process has been understood since the time of Francis Bacon to be a centerpiece of inductive reasoning.[23] The Method of Agreement is supported by the argument that whatever can be eliminated does not cause the phenomenon of interest, *a*. The Direct Method of Difference is supported by the argument that whatever cannot be eliminated does cause the phenomenon. Because both methods are based on the process of elimination, they are deterministic in nature. For if we observed even one case where effect *a* occurred without the presence of antecedent *A*, we would eliminate antecedent *A* from causal consideration.

Mill asserts that the Direct Method of Difference is commonly used in experimental science while the Method of Agreement, which is substantially weaker, is employed when experimentation is impossible.[24] The Direct Method of Difference is Mill's attempt to describe the inductive logic of experimental design. It takes into account two of the key features of experimental design: (1) the presence of a manipulation and (2) a comparison between two states of the world.[25] The method also incorporates the notion of a relative causal effect. The effect of antecedent *A* is measured in relation to the effect observed in the most similar world without *A*. The two states of the world we are considering only differ in the presence or absence of *A*.

The Direct Method of Difference accurately describes only a small subset of experiments. The method is too restrictive even if the relationship between antecedent *A* and effect *a* is deterministic. In particular, the control group *B*, *C* and the group with the intervention *A*, *B*, *C* need not be exactly alike (aside from the presence or absence of *A*). It would be fantastic if the two groups were exactly alike, but this is rarely possible to bring about. Some laboratory experiments are based on this strong assumption; but a more common assumption, one that brings in statistical concerns, is that observations in both groups

are balanced before our intervention. In other words, before we apply the treatment, the distributions of both observed and unobserved variables in both groups are presumed to be equal. For example, if group *A* is the southern states in the United States and group *B* is the northern states, the two groups are not balanced. The distribution of a long list of variables is different between the groups.

Random assignment of treatment ensures, if the sample is large and if other assumptions are met, that the control and treatment groups are balanced even on unobserved variables.[26] Random assignment also guarantees that the treatment is uncorrelated with all baseline variables,[27] whether we can observe them or not.[28] Thus, modern concepts of experimental design—because of their reliance on random assignment—sharply diverge from Mill's deterministic model.

If the balance assumption is satisfied, a modern experimenter estimates the relative causal effect by comparing the conditional probability of some outcome when the treatment is received with the outcome's conditional probability when the treatment is not received. In the canonical experimental setting, conditional probabilities can be directly interpreted as causal effects.

Complications arise when randomization of treatment is not possible. With observational data (which are found in nature, not as a product of experimental manipulation), many obstacles may prevent conditional probabilities from being directly interpreted as estimates of causal effects. Also problematic are experiments that prevent simple conditional probabilities from being interpreted as relative causal effects. (School voucher experiments are a good example of this phenomenon.[29]) But the most serious difficulties with observational data arise when neither manipulation nor balance is present.[30]

A primary reason that case study researchers find deterministic methods appealing is the power of the methods. For example, Mill's Direct Method of Difference can determine causality with only two observations. We assume that the antecedents *A*, *B*, *C* and *B*, *C* are exactly alike accept for the manipulation of *A*; we also assume deterministic causation as well as the absence of measurement error and interactions among antecedents. Once probabilistic factors are introduced, though, we need larger numbers of observations to make useful inferences. Unfortunately, because of the power of deterministic methods, social scientists with only a small number of observations are tempted to rely heavily on Mill's methods—particularly the Method of Agreement, which we have discussed, and the Indirect Method of Difference.

Indirect Method of Difference

> "If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance, the circumstance in which alone the two sets of instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon."   —John Stuart Mill[31]

This method arises by a "double employment of the Method of Agreement."[32] In a set of observations, if we note that effect

a is present and that the only antecedent in common is *A,* by the Method of Agreement we have evidence that *A* is the cause of *a.* Ideally, we then manipulate *A* to see if the effect *a* is present when the antecedent *A* is not. But when we cannot conduct such an experiment, we can instead use the Method of Agreement again. Suppose we find a set of observations in which neither the antecedent *A* nor the effect *a* occurs. We may now conclude, by use of the Indirect Method of Difference, that *A* is the cause of *a.* Thus, by twice using the Method of Agreement, we may hope to establish both the positive and negative instances that the Method of Difference requires.

However, this double use of the Method of Agreement is clearly inferior. The Indirect Method of Difference cannot fulfill the requirements of the Direct Method of Difference, for "the requisitions of the Method of Difference are not satisfied unless we can be quite sure either that the instances affirmative of *a* agree in no antecedents whatever but *A,* or that the instances negative of *a* agree in nothing but the negation of *A.*"[33] In other words, the Direct Method of Difference is the superior method because it entails a strong manipulation: we can remove the suspected cause, *A,* and then put it back at will, without disturbing the balance of what may lead to *a.* Thus, the only difference in the antecedents between the two observations is the presence or absence of *A.*

Many researchers are unclear about these distinctions between the Indirect and Direct Methods of Difference. They often simply state that they are using the Method of Difference when they are actually using only the Indirect Method of Difference. For example, Skocpol asserts that she is using both the Method of Agreement and the "more powerful" Method of Difference when she is at best using the weaker Method of Agreement twice.[34] Granted, Skocpol cannot use the Direct Method of Difference, since it is impossible to manipulate the factors of interest. But it is important to be clear about exactly which methods one employs.

n sum, scholars who claim to be using the Method of Agreement and the Method of Difference may actually be using the Indirect Method of Difference, the weaker sibling of the Direct Method of Difference. This weakness would not be of much concern if the phenomena we studied were simple. However, in the social sciences, we encounter serious causal complexities.

Mill's methods of inductive inference are valid only if the mapping between antecedents and effects is *unique* and *deterministic.*[35] These conditions allow neither for more than one cause of an effect nor for interactions among causes. In other words, if we are interested in effect *a,* we must assume a priori that only one possible cause exists for *a* and that when *a*'s cause is present—say, cause *A*—the effect *a* must *always* occur. These two conditions, uniqueness and determinism, also define the set of antecedents we are considering. The elements in the set of causes *A, B, C, D, E* must be able to occur independently of one another. The condition is not that antecedents must be independent in a probabilistic sense, but that any one of the antecedents can occur without requiring the presence or lack of any of the others. Otherwise, these rules cannot distinguish the possible effects of antecedents.[36]

The foregoing has a number of implications—most important, for deterministic methods such Mill's to work, there must be no measurement error. For even if there were a deterministic relationship between antecedent *A* and effect *a,* if we were able to measure either *A* or *a* only with some random measurement error, the resulting observed relationship would be probabilistic. We might, for instance, mistakenly think we have observed antecedent *A* (because of measurement error) in the absence of *a.* In such a situation, the process of elimination would lead us to conclude that *A* is not a cause of *a.*

To my knowledge, no modern social scientist argues that the conditions of uniqueness and lack of measurement error hold in the social sciences. However, the question of whether deterministic causation is plausible has a sizable literature.[37] Most of this discussion centers on whether deterministic relationships are possible—i.e., on the ontological status of deterministic causation.[38] Although such discussions can be fruitful, we need not decide the ontological issues in order to make empirical progress. This is fortunate, because the ontological issues are at best difficult to resolve and may be impossible to resolve. Even if we conceded that deterministic social associations exist, it is unclear how we would ever learn about them if there were multiple causes with complex interactions or if our measures were noisy. A case with multiple causes and complex interactions among deterministic associations would, to us, look probabilistic in the absence of a theory (and measurements) that accurately accounted for the complicated causal mechanisms.[39] There appears to be some agreement among qualitative and quantitative researchers that there is indeed complexity-induced probabilism.[40] Thus, I think it is more productive to focus on the practical issue of how we learn about causes—in other words, on the epistemological issues related to causality[41]—than on thorny philosophical questions regarding the ontological status of various notions of causality.[42]

Faced with multiple causes and interactions, what are we to do? There are two dominant responses. One relies on statistical tests that account for conditional probabilities and counterfactuals; the other, on detailed (usually formal) theories that make precise, distinct empirical predictions. The statistical approach is adopted by fields such as medicine that have access to large data sets and are able to conduct field experiments. In these fields experiments may be possible, but the available experimental manipulations are not strong enough to satisfy the requirements of the Direct Method of Difference. There are also fields in which researchers can conduct laboratory experiments with such strong manipulations and careful controls that a researcher may reasonably claim to have obtained exact balance and the practical absence of measurement error. These manipulations and controls allow generalizations of the Direct Method of Difference to be used. Deductive theories generally play a prominent role in such fields.[43] A number of theories in physics offer canonical examples.

These two responses are not mutually exclusive. Economics, for example, is a field that depends heavily on both formal

theories and statistical tests. Indeed, unless the proposed formal theories are nearly complete, there will always be a need to take random factors into account. And even the most ambitious formal modeler will no doubt concede that a complete deductive theory of politics is probably impossible. Given that our theories are weak, our causes complex, and our data noisy, we cannot avoid conditional probabilities. Thus, even researchers sympathetic to finding necessary or sufficient causes are often led to probability.[44]

Conditional Probability

Mill asks us to consider the situation in which we wish to ascertain the relationship between rain and any particular wind—say, the west wind. A particular wind will not always lead to rain, but the west wind may make rain more likely because of some causal relationship.[45]

How can we determine if rain and a particular wind are causally related? The simple answer is to observe whether rain occurs with one wind more frequently than with any other. But we need to take into account the baseline rate at which a given wind occurs. For example:

> In England, westerly winds blow about twice as great a portion of the year as easterly. If, therefore, it rains only twice as often with a westerly as with an easterly wind, we have no reason to infer that any law of nature is concerned in the coincidence. If it rains more than twice as often, we may be sure that some law is concerned; either there is some cause in nature which, in this climate, tends to produce both rain and a westerly wind, or a westerly wind has itself some tendency to produce rain.[46]

Formally, we are interested in the following inequality:

$$H : P(\text{rain} \mid \text{westerly wind}, \Omega) > P(\text{rain} \mid \textbf{not } \text{westerly wind}, \Omega),$$
$$(1)$$

where Ω is a set: of background conditions we consider necessary for a valid comparison. The probabilistic answer to our question is to compare the relevant conditional probabilities and to see if the difference between the two is significant.[42] In other words, our hypothesis is that the probability of rain given that there is a westerly wind (and given some background conditions we consider necessary) is larger than the probability of rain given that there is no westerly wind (and given the same background conditions as in the former case).

If we find that $P(\text{rain} \mid \text{westerly wind}, \Omega)$ is significantly larger than $P(\text{rain} \mid \textbf{not } \text{westerly wind}, \Omega)$, we would have some evidence of a causal relationship between westerly wind and rain. But many questions would remain unanswered. For example, we do not know whether the wind caused rain or vice versa. What is more disconcerting, there may be a common cause that results in both rain and the westerly wind; and without this common cause, the inequality above would be reversed. These caveats should alert us that there is much more to establishing causality than merely estimating some conditional probabilities. I will return to this issue in the penultimate section of this article.

Geddes's Critique of Skocpol

Geddes provides an excellent and wide-ranging discussion of case selection issues.[48] Unfortunately, her discussion of Skocpol's book *States and Social Revolutions* does not compare the relevant conditional probabilities, and this oversight results in a misleading conclusion.

Geddes's central critique is that Skocpol offers no contrasting cases when trying to establish her claim of a causal relationship between foreign threat and social revolution in her examination of the revolutions that occurred in France, Russia, and China. Geddes does point out that Skocpol provides contrasting cases—namely, England, Prussia/Germany, and Japan—when attempting to establish the importance of two causal variables: dominant classes having an independent economic base and peasants having autonomy.[49] But none are offered for the contention that

> developments within the international states system as such—especially defeats in wars or threats of invasion and struggles over colonial controls—have directly contributed to virtually all outbreaks of revolutionary crises.[50]

Geddes argues that many nonrevolutionary countries in the world have suffered foreign pressures at least as great as those suffered by the revolutionary countries Skocpol considers, but revolutions are nevertheless rare. Geddes points out that Skocpol first selects countries that have had revolutions and then notices that these countries have faced international threat—i.e., Skocpol has selected on the dependent variable. Such a research design ignores countries that are threatened but do not undergo revolution. In a proper (for instance, random) selection of cases, one could "determine whether revolutions occur more frequently in countries that have faced military threats or not."[51] Geddes acknowledges that obtaining and analyzing such a random sample is unrealistic. Even so, she argues that a rigorous test of Skocpol's thesis is possible because several Latin American countries have structural characteristics consistent with Skocpol's theory—such as village autonomy and economically independent dominant classes. Geddes considers Bolivia, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, and Peru, and asserts that although these countries have not been selected at random, their geographic location does not serve as a proxy for the dependent variable (revolution).

In short, Geddes claims that Skocpol has no variance in her dependent variable.[52] Scholars disagree on whether Skocpol seeks to establish only necessary conditions or necessary *and* sufficient conditions for social revolution.[53] As I note in the introduction of this article, Skocpol states that she relies on Mill's methods. But although she was clearly inspired by Mill, there remains considerable disagreement over the exact methods Skocpol uses. For example, Jack Goldstone argues that Mill's methods "are not used by comparative case-study analyses." He goes on to assert that it is "extremely unfortunate that . . . Theda Skocpol has identified her methods as Millian. . . . In fact, in many obvious ways, her methods depart sharply from Mill's canon."[54] Michael Burawoy goes even further and says that "applying [Mill's] principles would seem to falsify

Article | Quality Meets Quantity

Table 1
Relationship Between Defeat in War and Revolution in Latin America

	Revolution	No Revolution
Defeated and Invaded or Lost Territory	Bolivia: Defeated 1935, Revolution 1952	Peru, 1839 Bolivia, 1839 Mexico, 1848 Paraguay, 1869 Peru, 1883 Bolivia, 1883 Bolivia, 1903
Not Defeated within 20 Years	Mexico, 1910 Nicaragua, 1979	All Others

Source: Geddes 1990, 146.

[Skocpol's] theory."[55] He adds, though, that he nevertheless finds Skocpol's analysis compelling. William Sewell concurs with Burawoy and asserts that "it is remarkable, in view of the logical and empirical failure of [Skocpol's use of Mill's methods], that her analysis of social revolutions remains so powerful and convincing."[56]

Mahoney offers the most elaborate description of Skocpol's research design.[57] He concedes that when Skocpol applies Mill's methods to make nominal comparisons, the causal mechanisms under consideration must be deterministic. Yet, he points out, Skocpol also draws ordinal comparisons between her variables of interest, such as social revolution and foreign threat. She makes clear that foreign threat played a much larger role in the Russian revolution than in the French, even though she claims that both Russia and France faced serious foreign threats. Mahoney argues that these ordinal comparisons are "more consistent with the assumptions of statistical analysis" than are the nominal comparisons involved in Mill's methods.[58] This follows because when making ordinal comparisons, it is natural to examine how variables covary. And much of statistics is concerned with the analysis of covariance. In light of Mahoney's discussion, Skocpol's hypothesis of a causal relationship between foreign threat and revolution may be viewed as probabilistic. One can debate the soundness of this interpretation. But even if it were deemed unreasonable to interpret her theory as probabilistic, it would still be interesting to know whether a probabilistic relationship existed between foreign threat and revolution.

In Table 1,[59] you can see the relationship between foreign threat and revolution in the Latin American cases that Geddes considers relevant to Skocpol's theory. To demonstrate that Skocpol admits too many cases to the category of "foreign threat," Geddes claims that late-eighteenth-century France, Skocpol's canonical example, was "arguably the most powerful country in the world at the time" and was certainly less threatened than its neighbors.[60] Geddes's criterion for foreign threat is the loss of a war, accompanied by invasion or loss of territory. She does, however, use Skocpol's definition of revolution: a "rapid political and social structural change accompanied and, in part, caused by massive uprisings of the lower classes."[61] If a country faced a serious foreign threat and subsequently

(within 20 years) underwent revolution, Geddes classifies that country as a successful affirmative case for Skocpol's theory. Geddes lists seven cases of serious foreign threat that failed to result in revolution (foreign threat cannot be a sufficient condition), two revolutions that were not preceded by foreign threat (foreign threat cannot be a necessary condition), and one revolution that was consistent with Skocpol's argument. Based on this evidence, Geddes concludes that had Skocpol "selected a broader range of cases to examine, rather than selecting three cases because of their placement on the dependent variable, she would have come to different conclusions."[62]

In general, I find Geddes's application of Skocpol's theory to Latin American countries to be both reasonable and sympathetic to Skocpol's hypothesis.[63] However, Geddes's data do not support Skocpol's hypothesis if we interpret the hypothesis to imply a deterministic relationship. But as discussed previously, we may legitimately wish to establish whether there is a probabilistic association between foreign threat and revolution. Geddes herself appears to be interested in determining this, even though the analysis she presents does not allow for such a relationship. After all, she gathered her data so as to ascertain "whether revolutions occur more frequently in countries that have faced military threats or not."[64]

In order to decide whether the data support a probabilistic association, we need to compare two conditional probabilities. Recalling our discussion of winds and rain (1), we are interested in the following probabilities:

$$P(\text{revolution} \mid \text{foreign threat}, \Omega), \qquad (2)$$

$$P(\text{revolution} \mid \textbf{no} \text{ foreign threat}, \Omega), \qquad (3)$$

where Ω is the set of background conditions we consider necessary for valid comparisons (such as village autonomy and dominant classes who are economically independent). The probabilistic version of Skocpol's hypothesis is that the probability of revolution given foreign threat (2) is greater than the probability of revolution given the absence of foreign threat (3). Geddes never makes this comparison, but her table offers us the data to do so. We may estimate the first conditional probability of interest (2) to be $\frac{1}{8}$. In other words, according to the table, one observation (Bolivia, 1952) of the eight that experienced serious foreign threat underwent revolution.

An estimate of the second conditional probability of interest, the probability of revolution given that there is no serious foreign threat (3), must still be obtained. However, it is not clear from Geddes's table how many countries are in the "No Revolution" / "Not Defeated within 20 Years" cell. She only labels them "all others." Nevertheless, any reasonable manner

of filling this cell will result in an estimate for (3), which is a very small proportion—a much smaller proportion than the one-in-eight estimate obtained for (2). For example, let's take an extremely conservative approach and assume that in this "all others" cell we shall only include countries that do not appear in the other three cells of the table. Countries may appear multiple times in the table (notice Bolivia). We are left with four countries: Ecuador, El Salvador, Guatemala, and Honduras. Let us assume further that every 20 years since independence during which neither a revolution nor a defeat in a foreign war occurred in a given country counts as one observation for the "No Revolution" / "Not Defeated within 20 Years" cell of the table. This 20-year window is consistent with Geddes's decision to allow for 20 years between foreign defeat and revolution. Considering only these four countries, we arrive at 684 such years and hence roughly 34 observations. Since we have 34.2 20-year blocks with neither foreign defeat nor revolution, our estimate of (3) is: $\frac{2}{34.2+2} = \frac{1}{18.1}$. This number, $\frac{1}{18.1}$, is much smaller than our estimate of (2), which is $\frac{1}{8}$.

Instead of only considering the four countries that do not appear anywhere else in the table, if we consider all of the countries in 20-year blocks (starting from the date of independence and ending in 1989) with neither a revolution nor a foreign defeat, we are left with roughly 67 observations (1,337 years). This yields an estimate for (3) of $\frac{2}{66.85+2} = \frac{1}{34.425}$. Obviously, if we count every year as an observation, our estimate of (3) becomes even smaller.

It is not obvious how to determine whether these estimated differences between (2) and (3) are statistically significant. What is the relevant statistical distribution—a sampling distribution? a Bayesian posterior distribution?—of revolutions and significant foreign threats? However one answers that question, Geddes is clearly incorrect when she asserts that her table offers evidence that contradicts Skocpol's conclusions. Indeed, depending on the distributions of the key variables, the table may offer support for Skocpol's substantive points.[65]

Since publishing *States and Social Revolutions*, Skocpol has argued that "comparative historical analyses proceed through logical juxtapositions of aspects of small numbers of cases. They attempt to identify invariant causal configurations that necessarily (rather than probably) combine to account for outcomes of interest."[66] It is thus understandable why Geddes and many others have interpreted Skocpol as being interested in finding a deterministic relationship. Such an endeavor is highly problematic both in this specific case, given Table 1, and in general.[67] But Skocpol's substantive claim regarding the relationship between foreign threat and revolution, if interpreted as probabilistic, is plausible.

Nothing in this article should be taken as disagreement with Geddes's critique of Skocpol's research design. There is broad consensus that selection on the dependent variable leads to serious biases in inferences when probabilistic associations are of interest. But there is no consensus about the problems caused by selection issues when testing for necessary or sufficient causation. This is because scholars do not agree on what information is relevant in such testing.[68] Indeed, some even reject the

logic of deterministic elimination when counterexamples are observed—the logic upon which Mill's methods are based. To reach this conclusion, they rely on a particular form of measurement error,[69] the concept of "probabilistic necessity,"[70] or the related concept of "almost necessary" conditions.[71]

These attempts to bridge the gap between deterministic theories of causality and notions of probability are interesting. Although it is outside the scope of this article to fully engage them, I will note that once we admit that measurement error and causal complexity are problems, it is unclear what benefit there is in assuming that the underlying (but unobservable) causal relationship is in fact deterministic. This is an untestable proposition and hence one that should not be relied upon. It would appear to be more fruitful and straightforward to rely instead fully on the apparatus of statistical causal inference.[72]

No matter what inference one makes based on Table 1, Geddes is correct in saying that this exercise does not constitute a definitive test of Skocpol's argument. As we have seen, many of the decisions leading to the construction of Table 1 are debatable. But even if we resolve these debates in favor of Table 1, my conditional probability estimates may not provide accurate information of the counterfactuals of interest—e.g., whether a given country undergoing revolution would have been less likely to undergo revolution if it had not, *ceteris paribus*, faced the foreign threat it did. Moving from estimating conditional probabilities to making judgments about counterfactuals we never observe is tricky business.

From Conditional Probabilities to Counterfactuals

Although conditional probability is at the heart of causal inference, by itself it is not enough to support such inferences. Underlying conditional probability is a notion of counterfactual inference. It is possible to have a causal theory that makes no reference to counterfactuals,[73] but counterfactual theories of causality are by far the norm, especially in statistics.[74] The Direct Method of Difference is motivated by a counterfactual notion: I would like to see what happens both with and without antecedent A. Ideally, when I use the Direct Method of Difference, I do not conjecture what would happen if A were absent. I remove A and actually see what happens. Implementation of the method obviously depends on a manipulation. However, although manipulation is an important component of experiment research, manipulations as precise as those called for by the Direct Method of Difference are not possible in the social sciences or in field experiments generally.

We have to depend on other means to obtain information about what would occur if A were present and if A were not. In many fields, a common alternative to the Direct Method of Difference is a randomized experiment. For example, we can contact Jane to prompt her to vote as part of a turnout study, or we can *not* contact her. But we cannot do both. If we contact her, we must estimate what would have happened if we had not contacted her, in order to determine what effect contacting Jane has on her behavior (whether she votes or not). We could seek to compare Jane's behavior with that of someone we did

not contact who is exactly like her. The reality, however, is that no one is exactly like Jane (aside from the treatment received). So instead, in a randomized experiment, we obtain a group of people (the larger the better), contacting a randomly chosen subset and assigning the remainder to the control group (not to be contacted). We then observe the difference in turnout rates between the two groups and attribute any differences to our treatment.

In principle, the process of random assignment results in the observed and unobserved baseline variables of the two groups being balanced.[75] In the simplest setup, individuals in both groups are supposed to be equally likely to receive the treatment; therefore, assignment of treatment will not be associated with anything that also affects one's propensity to vote. Even in an experiment, much can go wrong that requires statistical correction.[76] In an observational setting, unless something special is done, the treatment and nontreatment groups are almost never balanced because treatment, such as foreign threat, is not randomly assigned. Assignment to treatment or control is not the result of manipulation by the scientist.

In the case of Skocpol's work on social revolutions, we would like to know whether countries that faced foreign threat would be less likely to undergo revolution if they had not faced such a threat, and vice versa. It is possible to consider foreign threat the treatment and revolution the outcome of interest. Countries with weak states may be more likely to undergo revolution and also more likely to be attacked by foreign adversaries. In that case, the treatment group (countries that faced external threat) and the control group (those that did not) are not balanced. Thus, any inferences about the counterfactual of interest based on the estimated conditional probabilities in the previous section would be erroneous. How erroneous the inferences will be depends on how unbalanced the two groups are.

Aspects of the previous two paragraphs are well understood by political scientists, especially if we replace the term *unbalanced groups* with the nearly synonymous *confounding variables*, or *left out variables*. But the core counterfactual motivation is often forgotten. This situation may arise when quantitative scholars attempt to estimate partial effects.[77] On many occasions, researchers estimate a regression and interpret each of the regression coefficients as estimates of causal effects, holding all of the other variables in the model constant. For many in the late nineteenth and early twentieth centuries, this was the goal of using regression in the social sciences. The regression model was to give the social scientist the kind of control that the physicist obtained through precise formal theories and the biologist gained through experiments. Unfortunately, if one's covariates are correlated with one another (as they almost always are), interpreting regression coefficients to be estimates of partial causal effects is usually asking too much from the data. With correlated covariates, one variable (such as race) does not move independently of other covariates (such as income, education, and neighborhood). With such correlations, it is difficult to posit interesting counterfactuals of which a single regression coefficient is a good estimate.

A good example of these issues is offered by the literatures that developed in the aftermath of the 2000 U.S. presidential election. A number of scholars have tried to estimate the relationship between voters' race and uncounted ballots. Ballots are uncounted because they contain either undervotes (no votes) or overvotes (more than the legal number of votes).[78] If we were able to estimate a regression model, for instance, showing no relationship between the race of a voter and his or her probability of casting uncounted ballots when and only when controlling for a long list of covariates, it would be unclear what we had found. This uncertainty holds even if ecological and a host of other problems are pushed aside because such a regression model may not allow us to answer the counterfactual question of interest: if a black voter became white, would this increase or decrease his or her chance of casting an uncounted ballot? What does it mean to change a voter from black to white? Given the data, it is not plausible that such a change would have no implications for the individual's income, education, or neighborhood of residence. It is difficult to conceptualize a serious counterfactual for which this regression result is relevant. Before any regression is estimated, we know that if we measure enough variables well, the race variable itself in 2000 will be insignificant. But in a world where being black is highly correlated with socioeconomic variables, it is not clear what we learn about the causality of ballot problems from a showing that the race coefficient itself can be made insignificant.

No general solutions or methods ensure that the statistical quantities we estimate provide useful information about the counterfactuals of interest. The solution, which almost always relies on research design and statistical methods, depends on the precise research question under consideration. But all too often, the problem is ignored, and the regression coefficient itself is considered to be an estimate of the partial causal effect. In sum, estimates of conditional means and probabilities are an important component of establishing causal effects, but they are not enough. We must also establish the relationship between the counterfactuals of interest and the conditional probabilities we have managed to estimate.[79]

Discussion

This article has by no means offered a complete discussion of causality and what it takes to demonstrate a causal relationship. There is much more to this process than just conditional probabilities or even counterfactuals. For example, it is often important to find the causal mechanism at work—to understand the sequence of events leading from *A* to *a*. I agree with qualitative researchers that case studies are particularly helpful in learning about such mechanisms. Process tracing is often cited as being especially useful in this regard.[80] But insofar as many occurrences of a given process are not compared, process tracing does not directly provide information about the conditional probabilities estimated in order to demonstrate a causal relationship.

The importance of searching for causal mechanisms is often overestimated by political scientists, and this sometimes leads to an underestimate of the importance of comparing conditional probabilities. We do not need to have much or any

knowledge about mechanisms in order to know that a causal relationship exists. For instance, owing to rudimentary experiments, aspirin has been known to help with pain since Felix Hoffmann synthesized a stable form of acetylsalicylic acid in 1897. In fact, the bark and leaves of the willow tree (rich in the substance called salicin) have been known to help alleviate pain at least since the time of Hippocrates. But only in 1971 did John Vane discover aspirin's biological mechanism of action.[81] And even now, although we know how aspirin crosses the blood-brain barrier, we have little idea how the chemical changes caused by aspirin get translated into the conscious feeling of pain relief—after all, the mind-body problem has not been solved. All the same, though, no causal account can be considered complete without a causal mechanism being demonstrated or, at the very least, hypothesized.

In clinical medicine, case studies continue to contribute valuable knowledge even though large-*n* statistical research dominates. Although the coexistence of case studies and large-n studies is sometimes uneasy, as shown by the rise of outcomes research, it is nevertheless extremely fruitful; clinicians and scientists are more cooperative than their counterparts in political science.[82] One reason for this is that in clinical medicine, researchers reporting cases more readily acknowledge that the statistical framework provides information about when and where cases are useful.[83] Cases can be highly informative when our understanding of the phenomena of interest is very poor, because then we can learn a great deal from a few observations. And when our understanding is generally very good, a few cases that combine a set of circumstances previously believed not to exist— or, more realistically, previously believed to be highly unlikely— can alert us to overlooked phenomena. Some observations are more important than others, and there sometimes are "critical cases."[84] This point is not new to qualitative methodologists, because their discussion of the relative significance of cases contains an implicit (and all too rarely explicit) Bayesianism.[85] If one has only a few observations, it is more imperative than usual to pay careful attention, when selecting cases and deciding how informative they are, to the existing state of knowledge. In general, as our understanding of an issue improves, studying individual cases becomes less important.

Notes

1 Mill 1872, 298.
2 Mill 1872.
3 Cohen and Nagel 1934.
4 Przeworski and Teune 1970.
5 A conditional probability is the probability of an event given that another event has occurred. For example, the probability that the total of two dice will be greater than 10 given that the first die is a 4 is a conditional probability.
6 Needless to say, although Mill was familiar with the work of Pierre-Simon Laplace and other nineteenth-century statisticians, by today's standards, his understanding of estimation and hypothesis testing was simplistic, limited, and— especially in terms of estimation—often erroneous. He

did, however, understand that if we want to make valid empirical inferences, we need to obtain and compare conditional probabilities when there may be more than one possible cause of an effect or when the causal relationship is complicated by interaction effects.
7 Geddes 1990; Skocpol 1979.
8 Geddes 1990, Figure 10.
9 Skocpol 1979, 36.
10 Skocpol does not make clear that she is, at best, using the Indirect Method of Difference, which is, as we shall see, much weaker than the Direct Method of Difference.
11 Holland 1986.
12 Ecological inferences are about individual behavior, based on data of group behavior.
13 Robinson 1951.
14 Przeworski and Teune 1970.
15 Ragin et al. 1996.
16 For all page referencing in *A System of Logic,* I have used a reprint of the eighth edition, which was initially published in 1872. The eighth edition was the last printed in Mill's lifetime. Of all the editions, the eighth and the third were especially revised and supplemented with new material.
17 Mill 1872.
18 Lieberson 1991.
19 I do not review Mill's other two canons, the Method of Residues and the Method of Concomitant Variations, because they are not directly relevant to my discussion.
20 Mill 1872, 255.
21 Per Mill, I use the word *antecedent* to mean "possible cause." Neither Mill nor I intend to imply that events must be ordered in time to be causally related.
22 Mill 1872, 256.
23 Pledge 1939.
24 Mill 1872.
25 The requirement of a manipulation by the researcher has troubled many philosophers of science. However, the claim is not that causality requires a human manipulation—only that if we wish to measure the effect of a given antecedent, we gain much if we are able to manipulate the antecedent. For instance, we can then be confident that the antecedent caused the effect, and not the other way around. See Brady 2002.
26 Aside from having a large sample size, experiments also need to meet a number of other conditions. See Campbell and Stanley 1966 for an overview particularly relevant for the social sciences. An important problem in experiments dealing with human beings is the issue of compliance. Full compliance implies that every person assigned to treatment actually receives it and every person assigned to control does not. Fortunately, if noncompliance is an issue, there are a number of possible corrections that make few and reasonable assumptions. See Barnard et al. 2003.
27 Baseline variables are the variables observed before treatment is applied.

28 More formally, random assignment results in the treatment being stochastically independent of all baseline variables as long as the sample size is large and other assumptions are satisfied.

29 Barnard et al. (2003) discuss in detail a broken school voucher experiment and a correction using stratification.

30 In an experiment, much can go wrong (e.g., compliance and missing data problems), but the fact that there is a manipulation can be very helpful in correcting the problems. See Barnard et al. 2003. Corrections are more problematic in the absence of an experimental manipulation because additional assumptions are required.

31 Mill 1872, 259.

32 Mill 1872, 258.

33 Ibid., 259.

34 Skocpol 1979.

35 Mill 1872.

36 Mill's methods have additional limitations that are outside the scope of this discussion. For example, there is a set of conditions, call it z, that always exists but is unconnected with the phenomenon of interest. The star Sirius, for instance, is always present (but not always observable) whenever it rains in Boston. Is Sirius and its gravitational force causally related to rain in Boston? Significant issues arise from this question, but I do not have room to address them here.

37 See Waldner 2002 for an overview.

38 Ontology is the branch of philosophy concerned with the study of existence itself.

39 For example, see Little 1998, chapter 11.

40 Bennett 1999.

41 Epistemology is the branch of philosophy concerned with the theory of knowledge—in particular, the nature and derivation of knowledge, its scope, and the reliability of claims to knowledge.

42 For example, if we can accurately estimate the probability distribution of A causing a, does that mean that we can explain any particular occurrence of a? After surveying three prominent theories of probabilistic causality in the mid-1980s, Wesley Salmon noted that "the primary moral I drew was that causal concepts cannot be fully explicated in terms of statistical relationships; in addition . . . we need to appeal to causal processes and causal interactions." Salmon 1989, 168. I do not think these metaphysical issues ought to concern practicing scientists.

43 Mill places great importance on deduction in the three step process of "induction, ratiocination, and verification." Mill 1872, 304. But on the whole, although the term *ratiocinative* is in the title of Mill's treatise and even appears before the term *inductive,* Mill devotes little space to the issue of deductive reasoning.

44 For example, see Ragin 2000.

45 Since a particular wind will not always lead to rain, this implies, according to Mill, that "the connection, if it exists, cannot be an actual law." Mill 1872, 346. However, he concedes that rain may be connected with a particular wind through some kind of causation. The fact that Mill reserves the word *law* to refer to deterministic relationships need not detain us.

46 Ibid., 346–7.

47 Mill had almost no notion of formal hypothesis testing, for it was rigorously developed only after Mill had died. He knew that the hypothesis test must be done, but he did not know how to formally do it. See Mill 1872.

48 Geddes 1990.

49 Geddes 1990.

50 Skocpol 1979, 23.

51 Geddes 1990, 144.

52 There have been a variety of responses to this charge. David Collier and James Mahoney (1996) concede that such a selection of cases does not allow a researcher to analyze covariation. As they note, the no-variance problem is not exclusively an issue with the dependent variable, and studies that lack variance on an independent variable are obviously also unable to analyze covariation with that variable. Collier and Mahoney argue, however, that a no-variance research design may all the same allow for fruitful inferences. Indeed, it is still possible to apply Mill's Method of Agreement. I have already discussed the Method of Agreement and the problems associated with it; see also Collier 1995.

Some scholars, contrary to Geddes, assert that Skocpol does have variation in her dependent variable, even when she considers the relationship between foreign threat and revolution. See Mahoney 1999, Table 2; Collier and Mahoney 1996. My discussion does not depend on resolving this disagreement.

53 Geddes's (1990) analysis assumes that Skocpol's theory posits variables individually necessary and collectively sufficient for social revolution. Douglas Dion (1998), in contrast, argues that Skocpol is proposing conditions that are necessary but not sufficient for social revolution.

54 Goldstone 1997, 108-9.

55 Burawoy 1989, 768.

56 Sewell 1996, 260.

57 He argues that Skocpol employs Mill's methods but that she also uses ordinal comparisons and narrative. Mahoney 1999. For our purposes here, it is the ordinal comparisons that matter. In the conclusion of this article, I discuss the importance of the narrative and process-tracing aspects of Skocpol's research design.

58 Ibid., 1,164.

59 In the original article, Geddes 1990, this is Figure 10.

60 Ibid., 143.

61 Ibid., 145.

62 Ibid., 145.

63 But one objection is that "none of the Latin American countries analyzed by Geddes fits Skocpol's specification of the domain in which she believes the causal patterns identified in her book can be expected to operate." Collier and Mahoney 1996, 81. Skocpol does assert in her book that she is concerned with revolutions in wealthy,

politically ambitious agrarian states that have not experienced colonial domination. Moreover, she explicitly excludes two cases (Mexico 1910 and Bolivia 1952) that Geddes includes in her analysis. I agree with Geddes, however, that it is not clear why the domain of Skocpol's precise causal theory should be so restricted. I do not attempt to resolve Skocpol and Geddes's disagreement regarding parameters. The following discussion is of interest no matter who is right on this point.

Another set of objections to Geddes's analysis concerns her operationalization of concepts. For instance, Dion (1998) claims that Mexico 1910 and Nicaragua 1979 should be moved to the "No Revolution" / "Not Defeated within 20 Years" cell. If Dion is right, one cannot eliminate the possibility that foreign threat is a necessary condition for social revolution. His argument is based, in part, on the understanding that the presence of a large number of cases in the "No Revolution" / "Not Defeated within 20 Years" cell is irrelevant in terms of evaluating necessary causation. This assumption is inaccurate—see Seawright 2002a and Seawright 2002b for details.

I acknowledge that such disagreements with Geddes may be legitimate, but they cut both ways. Goldstone, for example, argues that France was relatively free of foreign threat but nevertheless underwent revolution. Based on this and other points of contention over Skocpol's analysis, Goldstone concludes that "the incidence of war is neither a necessary nor a sufficient answer to the question of the causes of state breakdown." Goldstone 1991, 20.

Since my main interest here is methodological, I set aside these substantive disagreements and accept both Skocpol's and Geddes's operationalizations.

64 Geddes 1990, 144.

65 Some researchers may be tempted to make the usual assumption that all of the observations are independent. Pearson's well-known x^2 test of independence is inappropriate for this data because of the small number of observed counts in some cells. A reliable Bayesian method shows that 93.82 percent of the posterior density is consistent with our estimate of (2) being larger than our estimate of (3). Geddes's original table ends in 1989 because her article was published in 1990. If the table is updated to the end of 2003, the only change is that the count in the "No Revolution" / "Not Defeated within 20 Years" cell becomes 73. The Bayesian method then shows that 94.61 percent of the posterior density is consistent with our estimate of (2) being larger than our estimate of (3). See Sekhon 2003 for details.

66 Skocpol 1984, 378.

67 Lieberson 1994.

68 Braumoeller and Goertz 2002; Clarke 2002; Seawright 2002a; Seawright 2002b.

69 Braumoeller and Goertz 2000.

70 Dion 1998.

71 Ragin 2000.

72 This article has some similarities with Jason Seawright's (2002a) discussion of how to test for necessary or sufficient causation. Seawright and I, however, have different goals. He assumes that one wants to test for necessary or sufficient causation, and then goes on to demonstrate that all four cells in Geddes's table contain relevant information for such tests, even the "No Revolution" / "Not Defeated within 20 Years" cell. Nothing in Seawright 2002a alters the conclusion that, based on Table 1, one is able to reject the hypothesis that foreign threat is a necessary and/or sufficient cause of revolution. But I argue that one should test for probabilistic causation in the social sciences. And there is no disagreement in the literature that for such tests all four cells of Table 1 are of interest.

73 See Dawid 2000 for an example and Brady 2002 for a general review of causal theories.

74 Holland 1986; Rubin 1990; Rubin 1978; Rubin 1974; Splawa-Neyman 1990 [1923].

75 This occurs with arbitrarily high probability as the sample size grows.

76 Gerber and Green 2000; Imai (forthcoming); Rubin 1974; Rubin 1978.

77 These are the effects a given antecedent has when all of the other variables are held constant.

78 See Herron and Sekhon 2003 and Herron and Sekhon (forthcoming) for a review of the literature and relevant empirical analyses.

79 Many other issues are important in examining the quality of the conditional probabilities we have estimated. A prominent example is how and when we can legitimately combine a given set of observations—a question that has long been central to statistics. (In fact, a standard objection to statistical analysis is that observations rather different from one another should *not* be combined.) The original purpose of least squares was to give astronomers a way of combining and weighting their discrepant observations in order to obtain better estimates of the locations and motions of celestial objects. (See Stigler 1986.) A large variety of techniques can help analysts decide when it is valid to combine observations. For example, see Bartels 1996; Mebane and Sekhon 2004. This is a subject that political scientists need to give more attention.

80 Process tracing is the enterprise of using narrative and other qualitative methods to determine the mechanisms by which a particular antecedent produces its effects. See George and McKeown 1985.

81 He was awarded the 1982 Nobel Prize for Medicine for his discovery.

82 Returning to the aspirin example, it is interesting to note that Lawrence Craven, a general practitioner, noticed in 1948 that the 400 men to whom he had prescribed aspirin did not suffer any heart attacks. But it was not until 1985 that the U.S. Food and Drug Administration

(FDA) first approved the use of aspirin for the purposes of reducing the risk of heart attack. The path from Craven's observation to the FDA's action required a large-scale randomized experiment.

83 Vandenbroucke 2001.

84 Eckstein 1975.

85 In this context, Bayesianism is a way of combining a priori information with the information in the data currently being examined. See George and McKeown 1985; McKeown 1999.

References

Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin. 2003. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association* 98:462, 299–323.

Bartels, Larry M. 1996. Pooling disparate observations. *American Journal of Political Science* 40:3, 905–42.

Bennett, Andrew. 1999. Causal inference in case studies: From Mill's methods to causal mechanisms. Paper presented at the annual meeting of the American Political Science Association. Atlanta, 2–5 September.

Brady, Henry. 2002. Models of causal inference: Going beyond the Neyman-Rubin-Holland Theory. Paper presented at the nineteenth annual Summer Political Methodology Meetings, Seattle, 18–20 July.

Braumoeller, Bear F., and Gary Goertz. 2000. The methodology of necessary conditions. *American Journal of Political Science* 44:4, 844–58.

———. 2002. Watching your posterior: Comments on Seawright. *Political Analysis* 10:2, 198–203.

Burawoy, Michael. 1989. Two methods in search of science: Skocpol versus Trotsky. *Theory and Society* 18:6, 759–805.

Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Clarke, Kevin A. 2002. The reverend and the ravens: Comment on Seawright. *Political Analysis* 10:2, 194–7.

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

Collier, David. 1995. Translating quantitative methods for qualitative researchers: The case of selection bias. *American Political Science Review* 89:2, 461–6.

Collier, David, and James Mahoney. 1996. Insights and pitfalls: Selection bias in qualitative research. *World Politics* 49:1, 56–91.

Dawid, A. Philip. 2000. Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association* 95:450, 407–48.

Dion, Douglas. 1998. Evidence and inference in the comparative case study. *Comparative Politics* 30:2, 127–46.

Eckstein, Harry. 1975. Case study and theory in political science. In *Handbook of Political Science*. Vol. 7 of *Strategies of Inquiry*, eds. Fred I. Greenstein and Nelson W. Polsby. Reading, Mass.: Addison-Wesley, 79–137.

Geddes, Barbara. 1990. How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis* 2, 131–50.

George, Alexander L., and Timothy J. McKeown. 1985. Case studies and theories of organizational decision-making. In *Advances in Information Processing in Organizations*, eds. Robert F. Coulam and Richard A. Smith. Greenwich, Conn.: JAI Press, 21–58.

Gerber, Alan S., and Donald P. Green. 2000. The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* 94:3, 653–63.

Goldstone, Jack A. 1991. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press.

———. 1997. Methodological issues in comparative macro-sociology. *Comparative Social Research* 16, 107–20.

Herron, Michael C., and Jasjeet S. Sekhon. 2003. Overvoting and representation: An examination of overvoted presidential ballots in Broward and Miami-Dade Counties. *Electoral Studies* 22:1, 21–47.

———. Forthcoming. Black candidates and black voters: Assessing the impact of candidate race on uncounted vote rates. *Journal of Politics*.

Holland, Paul W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81:396, 945–60.

Imai, Kosuke. Forthcoming. Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments. *American Political Science Review*.

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* 70:2, 307–20.

———. 1994. More on the uneasy case for using Mill-type methods in small-N comparative studies. *Social Forces* 72:4, 1,225–37.

Little, Daniel. 1998. *Microfoundations, Method, and Causation: On the Philosophy of the Social Sciences*. New Brunswick, N.J.: Transaction Publishers.

Mahoney, James. 1999. Nominal, ordinal, and narrative appraisal in macrocausal analysis. *American Journal of Sociology* 104:4, 1,154–96.

McKeown, Timothy J. 1999. Case studies and the statistical worldview: Review of King, Keohane, and Verba's *Designing Social Inquiry: Scientific Inference in Qualitative Research*. *International Organization* 51:1, 161–90.

Mebane, Walter R., Jr., and Jasjeet S. Sekhon. 2004. Robust estimation and outlier detection for overdispersed multinomial models of count data. *American Journal of Political Science* 48:2, 391–410.

Mill, John Stuart. 1872 [1843]. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. 8th ed. London: Longmans, Green.

Pledge, Humphrey Thomas. 1939. *Science since 1500: A Short History of Mathematics, Physics, Chemistry, [and] Biology*. London: His Majesty's Stationery Office.

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Ragin, Charles C., Dirk Berg-Schlosser, and Gisèle de Meur. 1996. Political methodology: Qualitative methods. In *A New Handbook of Political Science*, eds. Robert E. Goodin and Hans-Dieter Klingemann. New York: Oxford University Press, 749–68.

Robinson, W. S. 1951. The logical structure of analytic induction. *American Sociological Review* 16:6, 812–8.

Rubin, Donald B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:5, 688–701.

———. 1978. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6:1, 34–58.

———. 1990. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* 5:4, 472–80.

Salmon, Wesley C. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Seawright, Jason. 2002a. Testing for necessary and/or sufficient causation: Which cases are relevant? *Political Analysis* 10:2, 178–93.

———. 2002b. What counts as evidence? Reply. *Political Analysis* 10:2, 204–7.

Sekhon, Jasjeet S. 2003. Making inferences from 2×2 tables: The inadequacy of the Fisher Exact Test and a reliable Bayesian alternative. Working paper. Available at jsekhon.fas.harvard.edu/papers/SekhonTables.pdf. Accessed 15 December 2003.

Sewell, William H. 1996. Three temporalities: Toward an eventful sociology. In *The Historic Turn in the Human Sciences*, ed. Terrence J. McDonald. Ann Arbor: University of Michigan Press, 245–80.

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

———. 1984. Emerging agendas and recurrent strategies in historical sociology. In *Vision and Method in Historical Sociology*, ed. Theda Skocpol. New York: Cambridge University Press, 356–91.

Splawa-Neyman, Jerzy. 1990 [1923]. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Trans. D. M. Dabrowska and T. P. Speed. *Statistical Science* 5:4, 465–72.

Stigler, Stephen M. 1986. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press.

Vandenbroucke, Jan P. 2001. In defense of case reports and case series. *Annals of Internal Medicine* 134:4, 330–4.

Waldner, David. 2002. Anti anti-determinism: Or what happens when Schrodinger's cat and Lorenz's butterfly meet Laplace's demon in the study of political and economic development. Paper presented at the annual meeting of the American Political Science Association, Boston, 29 August–1 September.