

# Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals.\*

Jasjeet S. Sekhon<sup>†</sup>

September 30, 2003

Word Count  $\approx$  10493

Subfield: Political Methodology

\*Forthcoming, *Perspectives on Politics*. I thank Walter R. Mebane, Jr., Henry Brady, Bear Braumoeller, Shigeo Hirano, John Londregan, Gary King, Theda Skocpol, Suzanne M. Smith, Jonathan N. Wand, the Editors, and three anonymous reviewers for valuable comments and advice. All errors are my responsibility.

<sup>†</sup>Assistant Professor, Department of Government, Harvard University. 34 Kirkland Street, Cambridge, MA, 02138. [jasjeet\\_sekhon@harvard.edu](mailto:jasjeet_sekhon@harvard.edu), [HTTP://jsekhon.fas.harvard.edu/](http://jsekhon.fas.harvard.edu/), Voice: 617.496.2426, Fax: 617.496.5149.

## **Abstract**

Case study research methods often contrast sharply with statistical methods. Methods which only consider deterministic relationships should not be used because serious inferential errors can be avoided if the basic lessons of statistical inference are heeded. Of particular importance is the use of conditional probabilities to compare relevant counterfactuals. A number of case study methods collectively referred to as Mill's methods have been used by generations of social science researchers and they contrast sharply with statistical methods. A prominent example of work using these methods is Skocpol's *States and Social Revolutions*. Geddes' 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' data support Skocpol's hypothesis. But Skocpol failed to provide the data necessary to compare conditional probabilities. Also, for causal inference conditional probabilities are of interest only insofar as they provide information about relevant counterfactuals.

## Introduction

“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.”<sup>1</sup>

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. However, case study research methods should not be in opposition to statistical methods. Case study research methods often rely on the assumption that the relationships between the variables of interest are deterministic. This is unfortunate because 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 (canons) of inductive inference formalized by John Stuart Mill in his *A System of Logic*.<sup>2</sup>

Mill’s methods have greatly influenced generations of researchers in the social sciences.<sup>3</sup> For example, the “most similar” and the “most different” research designs, which are often used in comparative politics, are variants of Mill’s methods.<sup>4</sup> 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 the acknowledged originator of the methods clearly described many of their limitations.

Mill’s and related methods are only valid when the hypothesized relationship between the cause and effect of interest is *unique* and *deterministic*. These two conditions imply other conditions such as the absence of measurement error which would cease to make the hypothesized causal relationship deterministic. These assumptions are strict, and they strongly restrict the

applicability of the methods. When these methods of inductive inference are not applicable, conditional probabilities<sup>5</sup> should be used to compare the relevant counterfactuals.<sup>6</sup>

The importance of comparing the conditional probabilities of relevant counterfactuals is sometimes overlooked by even good political methodologists. For example, Barbara Geddes in an insightful and often assigned article on case selection problems in comparative politics overlooks this issue when discussing Theda Skocpol's *States and Social Revolutions*.<sup>7</sup> Skocpol in this book explores the causes of social revolutions and examines the revolutions which occurred in France, Russia, and China and a few countries where revolutions did not occur, England, Prussia/Germany, and Japan. Geddes presents a table which purports to seriously question Skocpol's claim of a causal relationship between foreign threat and social revolution.<sup>8</sup> Geddes' evidence is compelling if the only possible relationship between foreign threat and social revolution is deterministic—i.e., if foreign threat is either a necessary or sufficient cause of social revolution. Geddes never considers the possibility that the relationship may be probabilistic. This consideration is of some importance because Geddes' data actually support Skocpol's hypothesized relationship if the relationship is interpreted to be probabilistic. My discussion does not, however, in any way undermine Geddes' criticism of Skocpol's research design for selecting on the dependent variable. In particular, 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”<sup>9</sup> and that “[i]ts logic was explicitly laid out by John Stuart Mill in his *A System of Logic*.”<sup>10</sup> She goes further and states that she is using *both* the Method of Agreement and the more powerful Method of Difference.<sup>11</sup> As practiced, comparative historical analysis may or may not have a distinguished pedigree, but as I shall demonstrate its pedigree cannot be traced to Mill.

For these methods to lead to valid inferences 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. 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.<sup>12</sup> But conditional probabilities are rarely of direct interest. When making causal inferences, we use conditional probabilities to learn about counterfactuals of interest—e.g., would social revolution have been less likely in Russia if Russia had not faced the foreign pressures it did? One has to be careful to establish the relationship between the counterfactuals of interest and the conditional probabilities one has managed to estimate. Researchers too often forget that this relationship must be established and rely instead on the statistical model itself.

In the next section I outline Mill's methods and show the serious limitations of Mill's canons and the need to formally compare conditional probabilities in all but the most limited of situations. I then discuss Geddes' critique of Skocpol and my elaborations and corrections. In the penultimate section I discuss how difficult it is to establish a relationship between the counterfactuals of interest and the conditional probabilities one has managed to estimate.

## **Mill's Methods and Conditional Probabilities**

### **The 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, Robinson, who is well known in political science for his work on the ecological inference problem,<sup>13</sup> also criticized the use of Mill-type methods of analytic induction in the social sciences.<sup>14</sup> Robinson's critique 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. Many other critics will be encountered in the course of our discussion.

Przeworski and Teune, in an influential book, advocate the use of what they call the “most similar” design and the “most different” design.<sup>15</sup> 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 Method of Difference. Although the Przeworski and Teune volume is over 30 years old, their argument continues to be influential. For example, Ragin, Berg-Schlosser, and de Meur in a recent review of qualitative methods make direct supportive references to both Mill’s methods and Przeworski and Teune’s formulations.<sup>16</sup>

Mill described his views on scientific investigations in *A System of Logic Ratiocinative and Inductive*, first published in 1843.<sup>17</sup> In an often cited chapter (bk. III, ch. 8), Mill formulates five guiding methods of induction: the Method of Agreement, the Method of Difference, the Double Method of Agreement and Difference (also known as the Indirect Method of Difference), the Method of Residues, and the Method of Concomitant Variations. These methods are often counted to be only four because the Double Method of Agreement and Difference may be considered to be just a derivative of the first two methods. This is a mistake because it obscures the tremendous difference between the combined method or what Mill calls the Indirect Method of Difference and the Direct Method of Difference.<sup>18</sup> Both the Method of Agreement and the Indirect Method of Difference, which is actually the Method of Agreement applied twice, are limited and require the machinery of probability in order to take chance into account when considering cases where the number of causes may be greater than one or where there may be interactions between the causes.<sup>19</sup> Other factors not well explored by Mill, such as measurement error, lead to the same conclusion.<sup>20</sup> The Direct Method of Difference is almost entirely limited to the experimental setting. And even in the case of the Direct Method of Difference, chance must be taken into account in the presence of measurement error or if there are interactions between causes which lead to probabilistic relationships between a cause,  $A$ , and its effect,  $a$ .

Next, we review Mill’s first three canons and show the importance of taking chance into account and comparing conditional probabilities when chance variations cannot be ignored.

## First Cannon: 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.”<sup>21</sup>

Assume that the *possible* causes, i.e., antecedents, under consideration are denoted by  $A, B, C, D, E$ , and the effect we are interested in is denoted by  $a$ .<sup>22</sup> An antecedent may be comprised of more than one constituent event or condition. 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,  $A$  may be defined as the presence of both permanganate ion and oxalic acid.

Let us further assume that we observe two observations and in the first we observe the antecedents  $A, B, C$ , and in the second we observe the antecedents  $A, D, E$ . If we also observe the effect,  $a$ , in both cases, we would say, following Mill’s Method of Agreement, that  $A$  is the cause of  $a$ . We conclude this because  $A$  was the only antecedent which occurred in both observations—i.e., the observations agree on the presence of antecedent  $A$ . This method has eliminated antecedents  $B, C, D, E$  as possible causes of  $a$ . Using this method, we endeavor to obtain observations which agree in the effect,  $a$ , and the supposed cause,  $A$ , but differ in the presence of other antecedents.

## Second Cannon: 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.”<sup>23</sup>

In the Method of Difference we require, contrary to the Method of Agreement, observa-

tions resembling one another in every other respect, but differing in the presence or absence of the antecedent we conjecture to be the true cause of  $a$ . If our object is to discover the effects of an antecedent  $A$ , we must introduce  $A$  into some set of circumstances we consider relevant, such as  $B, C$ , and having noted the effects produced, compare them with the effects of the remaining circumstances  $B, C$ , 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 of these methods are based on a process of elimination. This process has been understood since Francis Bacon to be a centerpiece of inductive reasoning.<sup>24</sup> The Method of Agreement is supported by the argument that whatever can be eliminated is not connected with the phenomenon of interest,  $a$ . The Method of Difference is supported by the argument that whatever cannot be eliminated is connected with the phenomenon by a law. Because both methods are based on the process of elimination, they are deterministic in nature. For if even one case is observed where effect  $a$  occurs without the presence of antecedent  $A$ , we would eliminate antecedent  $A$  from causal consideration.

Mill asserts that the Method of Difference is commonly used in experimental science while the Method of Agreement, which is substantially weaker, is employed when experimentation is impossible.<sup>25</sup> The Method of Difference is Mill's attempt to describe the inductive logic of experimental design. And the method takes into account two of the key features of experimental design, the first being the presence of a manipulation (treatment) and the second a comparison between two states of the world which are in Mill's case exactly alike aside from the presence of the antecedent of interest.<sup>26</sup> The method also incorporates the notion of a relative causal effect. The effect of antecedent  $A$  is measured relative 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 Method of Difference only accurately describes a small subset of experiments. The method is too restrictive even if the relationship between the antecedent  $A$  and effect  $a$  were to be deterministic. Today we would say that 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 such a situation is not only extremely difficult to find but also not necessary. Some laboratory experiments are based on this strong assumption, but a more common assumption, and one which brings in statistical concerns, is that observations in both groups are *balanced* before our intervention. That is, before we apply the treatment, the distributions of both observed and unobserved variables in both groups are 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 two 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.<sup>27</sup> Random assignment ensures that the treatment is uncorrelated with all baseline variables<sup>28</sup> whether we can observe them or not.<sup>29</sup>

Because of its reliance on random assignment, modern concepts of experimental design sharply diverge from Mill's deterministic model. The two groups are not exactly alike in baseline characteristics (as they would have to be in a deterministic setup), but, instead, their baseline characteristics have the same distribution. And consequently the baseline variables are uncorrelated with whether a particular unit received treatment or not.

When the balance assumption is satisfied, a modern experimenter estimates the relative causal effect by comparing the conditional probability of some outcome given the treatment minus the conditional probability of the outcome given that the treatment was not received. In the canonical experimental setting, conditional probabilities can be directly interpreted as causal effects.

In the penultimate section of this article, I discuss the complications which arise in using conditional probabilities to make causal inferences when randomization of treatment is not possible. With observational data (i.e., data found in nature and not a product of experimental manipulation), many complications arise which prevent conditional probabilities

from being directly interpreted as estimates of causal effects. Problems also often arise with experiments which prevent the simple conditional probabilities from being interpreted as relative causal effects. School voucher experiments are a good example.<sup>30</sup> But the problems are more serious with observational data where neither a manipulation nor balance are present.<sup>31</sup>

One of the continuing appeals of deterministic methods for case study researchers is the power of the methods. For example, Mill’s Method of Difference can determine causality with only two observations. This power can only be obtained by assuming that the observation with the antecedent of interest,  $A, B, C$  and the one without,  $B, C$  are exactly alike except for the manipulation of  $A$ , and by assuming deterministic causation and the absence of measurement error and interactions among antecedents. This power makes deterministic methods alluring for case study researchers, who generally don’t have many observations. Once probabilistic factors are introduced, larger numbers of observations are required to make useful inferences. Because of the power of deterministic methods, social scientists with a small number of observations are tempted to rely on Mill’s methods. Because these researchers cannot conduct experiments, they largely rely on the Method of Agreement, which we have discussed, and Mill’s third canon.

### **Third Canon: 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.”<sup>32</sup>

This method arises by a “double employment of the Method of Agreement”.<sup>33</sup> If we observe a set of observations in all of which we observe  $a$  and note that they have no antecedent in common but  $A$ , by the Method of Agreement we have evidence that  $A$  is the

cause of the effect  $a$ . Ideally, we would then perform an experiment where we manipulate  $A$  to see if the effect  $a$  is present when the antecedent  $A$  is absent. When we cannot conduct such an experiment, we can instead use the Method of Agreement again. Suppose, we can find another set of observations in which neither the antecedent  $A$  nor the effect  $a$  occur. 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 instance which 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  $A$ ”.<sup>34</sup> In other words, the Direct Method of Difference is the superior method because it entails a strong manipulation: we manipulate the antecedents so that we can remove the suspected cause,  $A$ , and then put it back at will, without disturbing the balance of what may lead to  $a$ . And this manipulation ensures that the only difference in the antecedents between the two observations is the presence of  $A$  or its lack.

Researchers are often unclear about these distinctions between the indirect and direct methods of difference. They often simply state they are using the Method of Difference when they are actually only using the Indirect Method of Difference. For example, Skocpol states that she is using both the Method of Agreement and the “more powerful” Method of Difference when she is only using at best the weaker Method of Agreement twice.<sup>35</sup> It is understandable that Skocpol is not able to use the Direct Method of Difference since it would be impossible to manipulate the factors of interest. But it is important to be clear about exactly which method one is using.

Mill discusses two other canons: Method of Residues (Fourth Canon); and the Method of Concomitant Variations (Fifth Canon). We do not review these canons because they are not directly relevant to our discussion.

We have so far outlined the three methods of Mill with which we are concerned. We have also shown that when scholars such as Skocpol assert that they are using the Method of Agreement and the Method of Difference,<sup>36</sup> they are actually using the Indirect Method of Difference, and that this is indeed 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*.<sup>37</sup> These conditions allow neither for more than one cause for an effect nor for interactions between 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. In fact, these two conditions, of uniqueness and determinism, define the set of antecedents we are considering. This implies, for example, that the elements in the set of causes  $A, B, C, D, E$  must be able to occur independently of each other. The condition isn't that antecedents must be independent in the probabilistic sense of the word, but that any one of the antecedents can occur without necessitating the presence or lack thereof of any of the other antecedents. Otherwise, the possible effects of antecedents are impossible to distinguish by these rules.<sup>38</sup> Generalizations of Mill's methods also suffer from these limitations.<sup>39</sup>

The foregoing has a number of implications the most important of which is that for deterministic methods such as 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 stochastic error, the resulting observed relationship would be probabilistic. It would be probabilistic because it would be possible to observe a case in which we mistakenly think we have observed antecedent  $A$  (because of measurement error) while not observing  $a$ . 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.<sup>40</sup> Most of this discussion centers on whether deterministic relationships are possible—i.e., on the ontological status of deterministic causation.<sup>41</sup> 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 one concedes that deterministic social associations exist, it is unclear how we would ever learn about them if there are multiple causes with complex interactions or if our measures are noisy. The case of multiple causes and complex interactions among deterministic associations would, to us, look probabilistic in the absence of a theory (and measurements) which accurately accounted for the complicated causal mechanisms.<sup>42</sup> There appears to be some agreement among qualitative and quantitative researchers that there is “complexity-induced probabilism.”<sup>43</sup> Thus, I think it is more fruitful to focus instead on the practical issue of how we learn about causes—i.e., on the epistemological issues related to causality.<sup>44</sup> Focusing on epistemological issues also helps to avoid some thorny philosophical questions regarding the ontological status of probabilistic notions of causality.<sup>45</sup>

Faced with multiple causes and interactions what is one to do? There are two dominant responses. The first relies on detailed (usually formal) theories which make precise empirical predictions which distinguish the theories. Such theories are usually tested by laboratory experiments with such strong manipulations and careful controls that one may reasonably claim to have obtained exact balance and the practical absence of measurement error. Such manipulations and controls allow one to use generalizations of the Method of Difference. A large number of theories in physics offer canonical examples of this approach. Deduction plays a prominent role in this approach.<sup>46</sup>

The second response relies on conditional probabilities and counterfactuals. These responses are not mutually exclusive. Economics, for example, is a field which relies heavily on both formal theories and statistical empirical 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 data noisy, we cannot avoid conditional probabilities. Even researchers sympathetic to finding necessary or sufficient causes are often led to probability given these problems.<sup>47</sup>

## 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.<sup>48</sup>

How would we determine if rain and a particular wind are related in some causal fashion? The simple answer would be 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.”<sup>49</sup>

Formally, we are interested in the following inequality:

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

where  $\Omega$  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.<sup>50</sup> In 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).

If we find that  $P(\text{rain}|\text{westerly wind}, \Omega)$  is significantly larger than  $P(\text{rain}|\text{not westerly wind}, \Omega)$ , we would have some evidence that there is some causal relationship between westerly wind and rain. But many questions would remain unanswered. For example, we would not know whether the wind caused rain or if rain caused westerly wind. More disconcertingly, there could be a common cause which resulted in both rain and the westerly wind and if not for this common cause, the inequality above would flip around. 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.

## Conditional Probabilities and Geddes' Critique

Geddes in an often cited article provides an excellent and wide-ranging discussion of case selection issues.<sup>51</sup> Unfortunately, in her discussion of Skocpol's *States and Social Revolutions* Geddes does not compare the relevant conditional probabilities even though she should, and this oversight leads to a misleading substantive conclusion.

Geddes' central critique of Skocpol's *States and Social Revolutions* is that Skocpol offers no contrasting cases when trying to establish her claim of a causal relationship between foreign threat and social revolution. Skocpol examines the revolutions which occurred in France, Russia, and China. Geddes concedes that Skocpol does offer contrasting cases for *some* of her central claims, namely England, Prussia/Germany, and Japan. Skocpol offers such cases when she attempts to establish the importance of two causal variables: domi-

nant classes having an independent economic base and the existence of peasant autonomy.<sup>52</sup> However, no contrasted cases 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.”<sup>53</sup>

Geddes argues that many 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 argues that Skocpol has selected countries which have had revolutions and has then noticed that these countries faced international threat—i.e., Skocpol has selected on the dependent variable. Such a research design ignores countries which are threatened, but do not undergo revolution. If one had a proper selection of cases (say random) one could then “determine whether revolutions occur more frequently in countries that have faced military threats or not”.<sup>54</sup> Geddes acknowledges that obtaining and analyzing such a random sample is unrealistic. Nevertheless, 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 dominant classes who are economically independent. Geddes considers Bolivia, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, and Peru. Geddes 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 (social revolution). Collier and Mahoney concede that such a selection of cases does not allow a researcher to analyze covariation.<sup>55</sup> They argue, however, that a no-variance research design may still allow for fruitful inferences. Indeed, it is still possible to apply Mill’s Method of Agreement.<sup>56</sup> We have already discussed the Method of Agreement and the problems associated with it.

Some scholars, contrary to Geddes, assert that Skocpol does have variation in her depen-



dent variable even when she considers the relationship between foreign threat and revolution.<sup>57</sup> My discussion in no way depends on resolving this disagreement.

Scholars also disagree if Skocpol is interested in establishing just necessary conditions or necessary and sufficient conditions for social revolution. Geddes' analysis<sup>58</sup> proceeds with the assumption that Skocpol's theory posits variables which are individually necessary and collectively sufficient for social revolution.<sup>59</sup> Dion, however, argues that Skocpol is proposing a set of conditions which are necessary but not sufficient for social revolution.<sup>60</sup>

As I note in the introduction, Skocpol clearly states that she is relying on Mill's methods. Although it is clear that Skocpol was inspired by Mill's methods, there is considerable disagreement over the exact methods Skocpol does use. For example, Jack Goldstone argues that Mill's methods "are not used by comparative case-study analyses".<sup>61</sup> 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".<sup>62</sup> Michael Burawoy agrees with Goldstone that Skocpol does not actually use Mill's methods.<sup>63</sup> In fact, Burawoy argues that "applying those [i.e., Mill's] principles would seem to falsify her [Skocpol's] theory".<sup>64</sup> Burawoy goes on to say 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".<sup>65</sup>

James Mahoney offers the most elaborate description of Skocpol's research design. Mahoney argues that Skocpol uses Mill's methods but that she also uses ordinal comparisons and narrative. For our current purposes, it is the ordinal comparisons which Mahoney discusses which are of interest.<sup>66</sup> Mahoney concedes that when Skocpol uses Mill's methods (to make nominal comparisons), the causal mechanisms under consideration must be deterministic. Skocpol, however, goes on to also make ordinal comparisons between her variables of interest, such as social revolution and foreign threat.<sup>67</sup> Mahoney argues that Skocpol's ordinal comparisons make clear that foreign threat played a much larger role in the Russian

revolution than the French, even though both Russia and France have the same scores on all of the variables if only dichotomous values are considered. Mahoney argues that these ordinal comparisons are “more consistent with the assumptions of statistical analysis” than the nominal comparisons involved in Mill’s methods.<sup>68</sup> This follows because when making ordinal comparisons, covariation is a central concept.

Given Mahoney’s discussion, it appears to be plausible to argue that Skocpol’s hypothesis of a causal relationship between foreign threat and revolution *may be* interpreted to be probabilistic. Whether this is reasonable or not appears to be a matter of debate in the literature. The relationship between foreign threat and revolution is, however, of interest even aside from any connection with Skocpol’s theory. Even if it were unreasonable to interpret Skocpol’s theory in a probabilistic manner, it would still be interesting to know if there indeed exists a probabilistic relationship between foreign threat and revolution.

Based on her sample of Latin American countries, Geddes obtains Table 1.<sup>69</sup> This table shows the relationship between foreign threats and revolutions in the Latin American cases that Geddes considers to be relevant to Skocpol’s theory. Geddes uses a higher level of foreign threat than Skocpol because Geddes claims that using late eighteenth-century France as the canonical example would admit too many cases to the foreign threat column. Geddes argues that France in the eighteenth-century was “arguably the most powerful country in the world at the time” and was certainly less threatened than its neighbors.<sup>70</sup> Geddes’ criterion for foreign threat is a loss of a war accompanied by invasion or loss of territory. Geddes does, however, directly borrow Skocpol’s definition of revolution. Thus, Geddes defines a revolution as a “rapid political and social structural change accompanied and, in part, caused by massive uprisings of the lower classes”.<sup>71</sup> If a revolution occurs within 20 years of a foreign threat, Geddes classifies a country as a successful affirmative case for Skocpol’s theory—i.e., a country which faced a serious foreign threat and which subsequently underwent revolution. I accept Geddes’ operationalization of Skocpol’s theory in Latin America without objection. Geddes’ decisions appear to be reasonable and sympathetic to Skocpol’s theory.

[\*\*\* Table 1 About here \*\*\*]

Examining Table 1, Geddes observes that there were seven cases of serious foreign threat which failed to result in revolution (foreign threat cannot be a sufficient condition), two revolutions which were not preceded by foreign threat (foreign threat cannot be a necessary condition) and one revolution which 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".<sup>72</sup>

One objection to Geddes' analysis 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".<sup>73</sup> Skocpol does assert in her book that she is concerned with revolutions in wealthy, politically ambitious agrarian states which have not experienced colonial domination. Moreover, Skocpol explicitly excludes two cases (Mexico 1910 and Bolivia 1952) which Geddes includes in her analysis. I agree with Geddes, however, that it is not clear given Skocpol's precise causal theory why the domain of the theory should be so restricted. Geddes affirmatively argues that these Latin American cases are within the domain of Skocpol's theory. I do not attempt to resolve this disagreement. The following discussion is of interest no matter who is right on this point.

Another set of objections to Geddes' analysis concerns disagreements over Geddes' operationalization of concepts. Dion argues that Mexico 1910 and Nicaragua 1979 should be moved to the "no revolution"/"not defeated" cell.<sup>74</sup> Hence, one could not eliminate the possibility that foreign threat is a necessary condition for social revolution.<sup>75</sup> I acknowledge that such disagreements may be legitimate, but such disagreements cut both ways. Goldstone, for example, argues that France was relatively free of foreign threat but it nevertheless underwent revolution.<sup>76</sup> Based on this and other examples of disagreements over operationalization with Skocpol, Goldstone concludes that "the incidence of war is neither a necessary nor a sufficient answer to the question of the causes of state breakdown".<sup>77</sup> Since

my main interest here is methodological and not substantive, I set aside these disagreements and accept Skocpol's operationalization of her data and Geddes' operationalization of her data.

If we leave aside the foregoing criticisms, Geddes' 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 determine if there is a probabilistic relationship between foreign threat and revolution. Geddes herself appears to be interested in determining if such a probabilistic association exists even though the analysis she presents does not allow for this. Geddes notes that she gathered her data so as to "determine whether revolutions occur more frequently in countries that have faced military threats or not".<sup>78</sup>

In order to determine if such a probabilistic association is supported by the data we need to compare two different conditional probabilities. Recalling our discussion of winds and rain (1), we are interested in the following probabilities:

$$P(\text{revolution}|\text{foreign threat}, \Omega), \tag{2}$$

$$P(\text{revolution}|\text{no foreign threat}, \Omega), \tag{3}$$

where  $\Omega$  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. An estimate of the first conditional probability of interest (2) may be obtained by looking at the first row in Table 1:  $\frac{1}{8}$ . In other words, according to the table, one country (Bolivia, 1952) of the eight who experienced serious foreign threat underwent revolution. If Skocpol's argument were that, given her background variables, foreign threat is a sufficient cause of revolution, the fraction one-in-eight would pose a serious problem. By a process of elimination (via the

Method of Agreement) we could determine that foreign threat is not a cause of revolution because in seven cases foreign threat did not lead to 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' table how many countries are in the "no revolution"/"not defeated" cell of the table. She only labels it "all others". Nevertheless, any reasonable manner in which to fill this cell will result in an estimate for (3) which is a very small proportion—i.e., a much smaller proportion than the one-in-eight estimate obtained for (2). For example, let's take an extremely conservative approach. Let's assume that in this "all others" cell we shall only consider countries which do not appear in the other three cells of the table. We are left with four countries: Ecuador, El Salvador, Guatemala and Honduras. Countries may appear multiple times in the table (notice Bolivia). Hence, let us assume 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" cell of the table. The ending date for this exercise was chosen to be 1989 because Geddes' article was published in 1990. This 20 year window is consistent with Geddes' decision to allow for a 20 year window 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 which are nowhere 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 (1337 years). This yields an estimate for (3) of  $\frac{2}{66.85 + 2} = \frac{1}{34.425}$ . This estimate is the one which is most consistent with the assumptions Geddes makes when constructing the other cells of Table 1.

It is not clear how to determine if these estimated differences between (2) and (3) are

statistically significant: what is the relevant statistical distribution—either a sampling distribution, a Bayesian posterior distribution or whatnot—of revolutions and significant foreign threats? However, it is clear that Geddes is incorrect when she asserts that Table 1 offers evidence which is contrary to Skocpol’s theory. Indeed, depending on the distributions of the key variables, the table may offer support for Skocpol’s substantive point.<sup>79</sup>

Subsequent to *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”.<sup>80</sup> Given this, it is understandable why Geddes and many others have interpreted Skocpol to mean that she was interested in finding a deterministic relationship. Such an endeavor is highly problematic both in this specific case given Table 1 and in general.<sup>81</sup> But Skocpol’s substantive claim regarding the relationship between foreign threat and revolution, if interpreted in a probabilistic fashion, is both plausible and of interest.

Nothing in this article should be taken to disagree with Geddes’ 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 there is no consensus regarding what information is relevant when testing for necessary and sufficient causation.<sup>82</sup> Indeed, some authors 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 either on a particular form of measurement error,<sup>83</sup> the concept of “probabilistic necessity”<sup>84</sup> or the related concept of “almost necessary” conditions.<sup>85</sup>

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 paper to fully engage them, I will note that once one admits that measurement error and causal complexity are

problems, it is unclear what benefit there is of assuming that the underlying (but unobservable) causal relationship is in fact deterministic. This is an untestable proposition, and hence one which 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.

This article has some similarities with Seawright’s discussion of how to test for necessary or sufficient causation.<sup>86</sup> Seawright and I, however, have different goals. In his article he assumes that one wants to test for necessary or sufficient causation. Seawright then goes on to demonstrate that all four cells in Table 1 contain relevant information for such tests including the “no revolution”/“not defeated” cell.<sup>87</sup> I, however, 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.

No matter what inference one makes based on Table 1, Geddes is correct in 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 the conditional probabilities we estimate to making judgments about counterfactuals we never observe is tricky business.

## From Conditional Probabilities to Counterfactuals

Although conditional probability is at the heart of inductive inference, by itself it isn’t enough. Underlying conditional probability is a notion of counterfactual inference. It is possible to have a causal theory that makes no reference to counterfactuals,<sup>88</sup> but counterfactual theories of causality are by far the norm, especially in statistics.<sup>89</sup> The Method of Difference is motivated by a counterfactual notion: I would like to see what happens both

with antecedent  $A$  and without  $A$ . When I use the Method of Difference, I don't 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. Although manipulation is an important component of experiment research, manipulations as precise as those entailed by the Method of Difference are not possible in the social sciences in particular and with field experiments in general.

We have to depend on other means to obtain information about both what would occur if  $A$  is present and if  $A$  is not. In many fields, a common alternative to the Method of Difference is a randomized experiment. For example, we could either contact Jane to prompt her to vote as part of a turnout study or we could not contact her. But we cannot observe what would happen if we both contacted Jane and if we did not contact Jane—i.e., we cannot observe Jane's behavior both with and without the treatment. If we contact Jane, in order to determine what effect this treatment had on Jane's behavior (i.e., whether she voted or not), we still have to obtain some estimate of the counterfactual in which we did not contact Jane. We could, for example, seek to compare Jane's behavior with someone exactly like Jane whom we did not contact. The reality, however, is that there is no one exactly like Jane (aside from the treatment) with whom we can compare Jane's turnout decision. Instead, in a randomized experiment we obtain a group of people (the larger the better) and we assign treatment to a randomly chosen subset (to contact) and we assign the remainder to the control group (not to be contacted). We then observe the difference in turnout rates between the two groups and we 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.<sup>90</sup> In the simplest setup, individuals in both groups are supposed to be equally likely to receive the treatment, and hence assignment to treatment will not be associated with anything which may also affect one's propensity to vote. Even in an experimental setup much can go wrong which requires statistical correction.<sup>91</sup> In an observational setting, unless something special is done, the treatment



and non-treatment groups are almost never balanced.

In the case of Skocpol’s work on social revolutions, we would like to know if countries which faced foreign threat would be less likely to undergo revolution if they had not faced such threats and vice-versa. It is possible to consider foreign threat the treatment and revolution the outcome of interest. It may be the case that countries with weak states are both more likely to undergo revolution and are more likely to be attacked by foreign adversaries. In that case the treatment group (those countries who faced external threat) and the control group (those countries who did not face external threat) 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 depends on how unbalanced the two groups are.

Aspects of the last two paragraphs are well understood by political scientists especially if we replace “unbalanced groups” with the nearly synonymous terms “confounding” or “left out variables.” But the core counterfactual motivation is often forgotten. This situation arises when quantitative scholars attempt to estimate partial effects.<sup>92</sup> On many occasions the researcher estimates a regression and interprets 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 19th and early 20th centuries, this was the goal of the use of regression in the social sciences. The regression model was to give the social scientist the control over data which the physicist obtained via precise formal theories and the biologist obtained via experiments. Unfortunately, if one’s covariates are correlated with each other (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 which developed in the after-

math of the 2000 Presidential election. A number of scholars try to estimate the relationship between the race of a voter and uncounted ballots. Ballots are uncounted either because the ballots contain no votes (undervotes) or overvotes (more than the legal number of votes).<sup>93</sup> If one were able to estimate a regression model, for example, which showed that there was no relationship between the race of a voter and her probability of casting uncounted ballots when and only when one controlled for a long list of covariates, it would be unclear what one has found. This uncertainty holds even if ecological and a host of other problems are pushed aside because such a regression model may not allow one to answer the counterfactual question of interest—i.e., “if a black voter became white, would this increase or decrease 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 changing a voter from black to white 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.

There are no general solutions or methods which 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. All too often the regression coefficient itself is considered to be an estimate of the partial causal effect. Estimates of conditional means and probabilities are an important component of establishing causal effects, but are not enough. One has to establish the relationship between the counterfactuals of interest and the conditional probabilities one has managed to estimate.

A large number of other issues are also important when one is examining the quality

of the conditional probabilities one has estimated. A prominent example is the extent to which one can combine a given collection of observations. The combining of observations which are actually rather different is one of the standard objections to statistical analysis. But the question of when and how one can legitimately combine observations is and has long been one of the central research questions in statistics. In fact, 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.<sup>94</sup> A large variety of techniques exist which can help the analyst make the decision of when it is legitimate to combine observations.<sup>95</sup> It is a subject that political scientists need to give more attention.

## Discussion

This article has by no means offered a complete discussion of causality and all one has to do in order to demonstrate a causal relationship. There is much more to this than just conditional probabilities and even counterfactuals. For example, it is often important to find the causal mechanism at work, in the sense of understanding the sequence of events which lead from  $A$  to  $a$ . And I agree with qualitative researchers that case studies are particularly helpful in learning about such mechanisms. Process tracing is often cited as being particularly useful in this regard.<sup>96</sup> But insofar as many occurrences of a given process are not compared, process tracing does not directly provide information about the conditional probabilities one has to estimate 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 example, by the use of 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) has been known to help alleviate pain at least since the time of Hippocrates. But the causal mechanism by which aspirin alleviates pain was a mystery until recently. Only in 1971 did John Vane discover aspirin's biological mechanism of action.<sup>97</sup> And even now, although we know how it crosses the blood-brain barrier, we have little idea how the chemical changes we can now demonstrate in the brain due to aspirin get translated into the conscious feeling of pain relief—after all, the mind-body problem has not been solved. But knowledge of causal mechanisms is important and useful and no causal account can be considered complete without a mechanism being demonstrated or at the very least hypothesized.

The search for causal mechanisms is probably especially useful when working with observational data. But it is still not necessary. In the case of the causal relationship between smoking and cancer, human experiments were not possible yet most (but not all) neutral researchers were convinced of the causal relationship well before the biological mechanisms were known.<sup>98</sup>

In clinical medicine case studies continue to contribute valuable knowledge even though large- $n$  statistical research dominates. Although the coexistence is sometimes uneasy, as noted by the rise of outcomes research, it is nevertheless extremely fruitful and more cooperative than the relationship in political science.<sup>99</sup> One reason for this is that in clinical medicine, researchers reporting cases more readily acknowledge that the statistical framework helps to provide information about when and where cases are informative.<sup>100</sup> 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. On the other hand, when our understanding is generally very good, a few cases which combine a set of circumstances that we believed could not exist or, more realistically, were believed to be highly unlikely can alert us to overlooked phenomena. Some observations are more important than others and there sometimes are “critical cases”.<sup>101</sup> This point is not new to qualitative methodologists

because there is an implicit (and all too rarely explicit) Bayesianism<sup>102</sup> in their discussion of the relative importance of cases.<sup>103</sup> When one only has a few observations, it is more important than usual to pay careful attention, when selecting cases and when deciding how informative they are, to the existing state of knowledge. In general, as our understanding of an issue improves, individual cases become less important.

## 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. 1996. Pooling Disparate Observations. *American Journal of Political Science* 40:3, 905–942.
- Bennett, Andrew. 1998. 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, GA.
- Brady, Henry. 2002. Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland Theory. Paper presented at the 19th Annual Summer Political Methodology Meetings. Seattle, WA.
- Braumoeller, Bear F. and Gary Goertz. 2000. The Methodology of Necessary Conditions. *American Journal of Political Science* 44:4, 844–858.
- Braumoeller, Bear F. and Gary Goertz. 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 Company.
- Clarke, Keven A. 2002. The Reverend and the Ravens: Comment on Seawright. *Political Analysis* 10:2, 194–197.

- Cohen, Morris and Ernest Nagel. 1934. *An Introduction to Logic and Scientific Method*. Harcourt, Brace and Company.
- Collier, David. 1995. Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias. *American Political Science Review* 89:2, 461–466.
- Collier, David and James Mahoney. 1996. Insights and Pitfalls: Selection Bias in Qualitative Research. *World Politics* 49:1, 56–91.
- Dawid, A. Phillip. 2000. Causal Inference without Counterfactuals (with Discussion). *Journal of the American Statistical Association* 95:450, 407–424.
- Dion, Douglas. 1998. Evidence and Inference in the Comparative Case Study. *Comparative Politics* 30:2, 127–146.
- Eckstein, Harry. 1975. Case Study and Theory in Political Science. In *Handbook of Political Science. Vol. 7. Strategies of Inquiry*, eds. Fred I. Greenstein and Nelson W. Polsby. Reading, MA: Addison-Wesley, 79–137.
- Fisher, Ronald A. 1958a. Cancer and Smoking. *Nature* 182 (August), 596.
- Fisher, Ronald A. 1958b. Lung Cancer and Cigarettes? *Nature* 182 (July), 108.
- Geddes, Barbara. 1990. How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2, 131–150.
- 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, CT: 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, 653–663.

- Goldstone, Jack A. 1991. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press.
- Goldstone, Jack A. 1997. Methodological Issues in Comparative Macrosociology. *Comparative Social Research* 16, 107–120.
- 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.
- Herron, Michael C. and Jasjeet S. Sekhon. 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, 945–960.
- Kosuke, Imai. 2003. Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments. Working Paper.
- 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, 307–320.
- Lieberson, Stanley. 1994. More On the Uneasy Case for Using Mill-Type Methods in Small-N Comparative Studies. *Social Forces* 72:4, 1225–1237.
- Little, Daniel. 1998. *Microfoundations, Method, and Causation*. New Brunswick, NJ: Transaction Publishers.
- Mahoney, James. 1999. Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis. *American Journal of Sociology* 104:4, 1154–1196.
- 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, 161–190.



- Mebane, Walter R., Jr. and Jasjeet S. Sekhon. forthcoming. Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data. *American Journal of Political Science*.
- Mill, John Stuart. 1872. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green and Co. 8th edition.
- Pledge, Humphry Thomas. 1939. *Science Since 1500: A Short History of Mathematics, Physics, Chemistry [and] Biology*. London: His Majesty's Stationery Office.
- Przeworski, A. and H. Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.
- 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, 749–768.
- Robinson, William S. 1951. The Logical Structure of Analytic Induction. *American Sociological Review* 16:6, 812–818.
- Rubin, Donald B. 1974. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. *Journal of Educational Psychology* 66:5, 688–701.
- Rubin, Donald B. 1978. Bayesian Inference for Causal Effects: The Role of Randomization. *Annals of Statistics* 6:1, 34–58.
- Rubin, Donald B. 1990. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science* 5:4, 472–480.
- 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, 178–193.
- Seawright, Jason. 2002b. What Counts as Evidence? Reply. *Political Analysis* 10:2, 204–207.
- Sekhon, Jasjeet S. 2003. Making Inferences from  $2 \times 2$  Tables: The Inadequacy of the Fisher Exact Test and a Reliable Bayesian Alternative. Working Paper.  
URL <http://jsekhon.fas.harvard.edu/papers/SekhonTables.pdf>
- 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, 245–280.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge, UK: Cambridge University Press.
- Skocpol, Theda. 1984. Emerging Agendas and Recurrent Strategies in Historical Sociology. In *Vision and Method in Historical Sociology*, ed. Theda Skocpol. New York: Cambridge University Press, 356–391.
- Splawa-Neyman, Jerzy. 1923 [1990]. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science* 5:4, 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed.
- Stigler, Stephen M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Vandenbroucke, Jan P. 2001. In Defense of Case Reports and Case Series. *Annals of Internal Medicine* 134:4, 330–334.
- Waldner, David. 2002. Anti Anti-Determinism: Or What Happens When Schrödinger’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, MA.

## Notes

<sup>1</sup>Mill 1872, 298.

<sup>2</sup>Mill 1872.

<sup>3</sup>Cohen and Nagel 1934.

<sup>4</sup>Przeworski and Teune 1970.

<sup>5</sup>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.

<sup>6</sup>Needless to say, although Mill was familiar with the work of Laplace and other 19th century statisticians, by today's standards his understanding of estimation and hypothesis testing was simplistic, limited and, especially when it comes to estimation, often erroneous. He did, however, understand that if one wants to make valid empirical inferences, one needs 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.

<sup>7</sup>Geddes 1990; Skocpol 1979.

<sup>8</sup>Geddes 1990, Figure 10.

<sup>9</sup>Skocpol 1979, 36.

<sup>10</sup>Skocpol 1979, 36.

<sup>11</sup>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 actual or direct Method of Difference.

<sup>12</sup>Holland 1986.

<sup>13</sup>Ecological inferences are inferences about individual behavior which are based on data of group behavior, called aggregate or ecological data.

<sup>14</sup>Robinson 1951.

<sup>15</sup>Przeworski and Teune 1970.

<sup>16</sup>Ragin et al. 1996.

<sup>17</sup>For all page referencing I have used a reprint of the eighth edition of *A System of Logic*

*Ratiocinative and Inductive*, first published in 1872. The eighth edition was the last printed in Mill’s lifetime. The eighth and third editions were especially revised and supplemented with new material.

<sup>18</sup>Mill 1872, 259.

<sup>19</sup>Mill 1872, 344.

<sup>20</sup>Lieberson 1991.

<sup>21</sup>Mill 1872, 255.

<sup>22</sup>Following Mill’s usage, my usage of the word “antecedent” is synonymous with “possible cause.” Neither Mill nor I intend to imply that events *must* be ordered in time to be causally related.

<sup>23</sup>Mill 1872, 256.

<sup>24</sup>Pledge 1939.

<sup>25</sup>Mill 1872, 256.

<sup>26</sup>The requirement of a manipulation by the researcher has troubled many philosophers of science. But the claim is not that causality requires a human manipulation, but 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 example, manipulation of the antecedent of interest allows us to be confident that the antecedent caused the effect and not the other way around—see Brady 2002.

<sup>27</sup>Aside from a large sample size, experiments need to also meet a number of other conditions. See Campbell and Stanley 1966 for an overview particularly relevant for the social sciences. An important problem with experiments dealing with human beings is the issue of compliance. Full compliance implies that every person assigned to treatment actually receives the treatment and every person assigned to control does not. Fortunately, if non-compliance is an issue, there are a number of possible corrections which make few and reasonable assumptions—see Barnard, Frangakis, Hill, and Rubin 2003.

<sup>28</sup>Baseline variables are the variables observed before treatment is applied.

<sup>29</sup>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.

<sup>30</sup>Barnard et al. 2003 discuss in detail a broken school voucher experiment and a correction using stratification.

<sup>31</sup>In an experiment much can go wrong (e.g., compliance and missing data problems), but the fact that there was a manipulation can be very helpful in correcting the problems—Barnard et al. 2003. Corrections are more problematic in the absence of an experimental manipulation because additional assumptions are required.

<sup>32</sup>Mill 1872, 259.

<sup>33</sup>Mill 1872, 258.

<sup>34</sup>Mill 1872, 259.

<sup>35</sup>Skocpol 1979, 36–37.

<sup>36</sup>Skocpol 1979, 37.

<sup>37</sup>Mill 1872, 285–299, 344–350.

<sup>38</sup>Mill’s methods have additional limitations which are outside the scope of this discussion. For example, there is a set of conditions, call it  $Z$ , which always exists but is unconnected with the phenomenon of interest. For example, the star Sirius is always present (but not always observable) whenever it rains in Boston. Is the star Sirius and its gravitational force causally related to rain in Boston? Significant issues arise from this question which I do not discuss.

<sup>39</sup>Little 1998, 221–223.

<sup>40</sup>See Waldner 2002 for an overview.

<sup>41</sup>Ontology is the branch of philosophy concerned with the study of existence itself.

<sup>42</sup>For example, Little 1998, ch. 11.

<sup>43</sup>Bennett 1998.

<sup>44</sup>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.

<sup>45</sup>For example, if one can accurately estimate the probability distribution of  $A$  causing  $a$ , does that mean that we can explain any particular occurrence of  $a$ ? Wesley Salmon after surveying three prominent theories of probabilistic causality in the mid-1980s noted that “the primary moral I drew was that causal concepts cannot be fully explicated in terms of statistical relationships; in addition, I concluded, 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.

<sup>46</sup>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.

<sup>47</sup>For example see Ragin 2000, 107–115.

<sup>48</sup>Since a particular wind will not always lead to rain Mill claims that this implies that “the connection, if it exists, cannot be an actual law” Mill 1872, 346. However, Mill still 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.

<sup>49</sup>Mill 1872, 346–347.

<sup>50</sup>Mill had almost no notion of formal hypothesis testing, for it was rigorously developed only after Mill had died—see Mill 1872, 350–360. Mill knew that the hypothesis test must be done, but he did not know how to formally do it.

<sup>51</sup>Geddes 1990.

<sup>52</sup>Geddes 1990, 141–142.

<sup>53</sup>Skocpol 1979, 23.

<sup>54</sup>Geddes 1990, 144.

<sup>55</sup>As Collier and Mahoney 1996, 72–75 note, the no-variance problem is not exclusively an issue with the dependent variable, and studies which lack variance on an independent variable are obviously also unable to analyze covariation with that variable.

<sup>56</sup>Also see Collier 1995, 464.

<sup>57</sup>Mahoney 1999, Table 2; Collier and Mahoney 1996, 80.

<sup>58</sup>Geddes 1990, 142–143.

<sup>59</sup>Dion 1998, 114, fn. 28.

<sup>60</sup>Dion 1998, 130.

<sup>61</sup>Goldstone 1997, 108.

<sup>62</sup>Goldstone 1997, 109.

<sup>63</sup>Burawoy 1989.

<sup>64</sup>Burawoy 1989, 768.

<sup>65</sup>Sewell 1996, 260.

<sup>66</sup>I discuss the importance of the narrative and process tracing aspects of Skocpol’s research design in the conclusion.

<sup>67</sup>Mahoney 1999, 1162, Table 2.

<sup>68</sup>Mahoney 1999, 1164.

<sup>69</sup>Figure 10 in the original article Geddes 1990, 146.

<sup>70</sup>Geddes 1990, 143.

<sup>71</sup>Geddes 1990, 145.

<sup>72</sup>Geddes 1990, 145.

<sup>73</sup>Collier and Mahoney 1996, 81.

<sup>74</sup>Dion 1998, 131.

<sup>75</sup>Dion’s 1998, 131 argument is based, in part, on the understanding that the presence of a large number of cases in the “no revolution”/“not defeated” cell is *irrelevant* when evaluating necessary causation. This is inaccurate—see Seawright 2002a; 2002b for details.

<sup>76</sup>Goldstone 1991.



<sup>77</sup>Goldstone 1991, 20.

<sup>78</sup>Geddes 1990, 144.

<sup>79</sup>Some researchers may be tempted to make the usual assumption that all of the observations are independent. Pearson’s well known  $\chi^2$  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% of the posterior density is consistent with our estimate of (2) being larger than our estimate of (3). The original table ended in 1989 because Geddes’ article was published in 1990. If the table is updated to the end of 2002, the only change is that the count in the “No Revolution”/“Not Defeated” cell becomes 73. The Bayesian method then shows that 94.61% of the posterior density is consistent with our estimate of (2) being larger than our estimate of (3). See Sekhon (2003) for details.

<sup>80</sup>Skocpol 1984, 378.

<sup>81</sup>Liebertson 1994, 1991.

<sup>82</sup>Braumoeller and Goertz 2002; Clarke 2002; Seawright 2002a,b.

<sup>83</sup>Braumoeller and Goertz 2000.

<sup>84</sup>Dion 1998.

<sup>85</sup>Ragin 2000.

<sup>86</sup>Seawright 2002a.

<sup>87</sup>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.

<sup>88</sup>See Dawid 2000 for an example and Brady 2002 for a general review of causal theories.

<sup>89</sup>Holland 1986; Rubin 1990, 1978, 1974; Splawa-Neyman 1923 [1990].

<sup>90</sup>This occurs with arbitrarily high probability as the sample size grows.

<sup>91</sup>Gerber and Green 2000; Kosuke 2003; Rubin 1978, 1974.

<sup>92</sup>A partial effect is the effect a given antecedent has on the outcome variable net of all the other antecedents—i.e., when all of the other variables “are held constant.”

<sup>93</sup>See Herron and Sekhon 2003 and forthcoming for a review of the literature and relevant

empirical analysis.

<sup>94</sup>Stigler 1986.

<sup>95</sup>For example, see Bartels 1996; Mebane and Sekhon forthcoming.

<sup>96</sup>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.

<sup>97</sup>He was awarded the 1982 Nobel Prize for Medicine for his discovery.

<sup>98</sup>R. A. Fisher, one of the fathers of modern statistics and the experimental method, was a notable exception. Without the manipulation offered by an experiment, he remained skeptical. He hypothesized that genetic factors could cause people to both smoke and get cancer, and hence there need not be any causal relationship between smoking and cancer. Fisher 1958b,a.

<sup>99</sup>Returning to the aspirin example, it is interesting to note that Lawrence Craven, a general practitioner, noticed in 1948 that the 400 men he had prescribed aspirin to 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.

<sup>100</sup>Vandenbroucke 2001.

<sup>101</sup>Eckstein 1975.

<sup>102</sup>In this context, Bayesianism is a way of combining *a priori* information with the information in the data currently being examined.

<sup>103</sup>George and McKeown 1985, 38; McKeown 1999.

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

Note: From Geddes 1990, 146, Figure 10.