How should we judge competing explanatory claims in social science research? How can we make inferences about which alternative explanations are more convincing, in what ways, and to what degree? Case study methods—especially methods of within-case analysis such as process tracing—are an indispensable part of the answer to these questions (George and Bennett 2005: chap. 10). This chapter offers an overview of process tracing as a tool for causal inference, focusing on the study of international relations, an area rich with examples of this approach. In contrast to the subsequent two chapters in this volume (chaps. 11 and 12), where Freedman and Brady analyze micro-level examples, the present chapter explores process tracing in macro studies.

This chapter uses three explanatory puzzles, about which scholars have advanced contending hypotheses, to illustrate how process tracing helps adjudicate among alternative explanations: (1) why and how the United
Kingdom and France resolved their competing imperial claims to the Upper Nile Valley without resorting to the use of force in the Fashoda crisis of 1898, an outcome that has been the subject of considerable research given its relevance to the inter-democratic peace hypothesis; (2) why in the middle of World War I, despite strong evidence that it was likely to be defeated, Germany expanded its war goals—for example, shifting to unrestricted submarine warfare—even though this risked (and in fact, resulted in) American entry into the conflict; and (3) why the Soviet Union did not intervene militarily in the Central European revolutions of 1989, in contrast to its military interventions in Hungary in 1956 and Czechoslovakia in 1968.

OVERVIEW OF PROCESS TRACING

Process tracing involves the examination of “diagnostic” pieces of evidence within a case that contribute to supporting or overturning alternative explanatory hypotheses. A central concern is with sequences and mechanisms in the unfolding of hypothesized causal processes. The researcher looks for the observable implications of hypothesized explanations, often examining evidence at a finer level of detail or a lower level of analysis than that initially posited in the relevant theory. The goal is to establish whether the events or processes within the case fit those predicted by alternative explanations.

This mode of analysis is closely analogous to a detective attempting to solve a crime by looking at clues and suspects and piecing together a convincing explanation, based on fine-grained evidence that bears on potential suspects’ means, motives, and opportunity to have committed the crime in question. It is also analogous to a doctor trying to diagnose an illness by taking in the details of a patient’s case history and symptoms and applying diagnostic tests that can, for example, distinguish between a viral and a bacterial infection (Gill, Sabin, and Schmid 2005).

Process tracing, which focuses on the diagnostic intervening steps in a hypothesized causal process, can provide inferential leverage on two problems that are difficult to address through statistical analysis alone. The first is the challenge of establishing causal direction: if X and Y are correlated, 2. A case may be understood as a temporally and spatially bounded instance of a specified phenomenon. Although process tracing focuses on events within a case, it can play a role in comparisons of cases. An analyst can use process tracing, for example, to assess whether a variable whose value differs in two most similar cases is related to the difference in their outcomes.

3. Process tracing is also used as a method of discovering hypotheses, a contribution illustrated below in Freedman’s contribution (chap. 11). However, that facet is not addressed in the present chapter.
did X cause Y, or did Y cause X? Careful process tracing focused on the sequencing of who knew what, when, and what they did in response, can help address this question. It might, for example, establish whether an arms race caused a war, or whether the anticipation of war caused an arms race.

A second challenge is that of potential spuriousness: If X and Y are correlated, is this because X caused Y, or is it because some third variable caused both X and Y? Here, process tracing can help establish whether there is a causal chain of steps connecting X to Y, and whether there is such evidence for other variables that may have caused both X and Y. There is no guarantee that researchers will include in their analyses the variable(s) that actually caused Y, but process tracing backward from observed outcomes to potential causes—as well as forward from hypothesized causes to subsequent outcomes—allows researchers to uncover variables they have not previously considered. This is similar to how a detective can work forward from suspects and backwards from clues about a crime. It is likewise consistent with David Freedman’s argument (chap. 11, this volume) that case expertise and substantive knowledge can play a key role in sorting out explanations—a claim that may for some readers appear counter-intuitive in light of Freedman’s disciplinary background as a mathematical statistician.

Critics have raised two critiques of process tracing: the “infinite regress” problem and the “degrees of freedom” problem. On the former, King, Keohane, and Verba suggest that the exceedingly fine-grained level of detail involved in process tracing can potentially lead to an infinite regress of studying “causal steps between any two links in the chain of causal mechanisms” (1994: 86). Others have worried that qualitative research on a small number of cases with a large number of variables suffers from a degrees of freedom problem. This form of indeterminacy afflicts statistical studies, given that the number of cases in a data set must be far greater than the number of variables in a model to test that model through frequentist statistics.

The answer to both critiques is that not all data are created equal. With process tracing, not all information is of equal probative value in discriminating between alternative explanations, and a researcher does not need to examine every line of evidence in equal detail. It is possible for one piece of evidence to strongly affirm one explanation and/or disconfirm others, while at the same time numerous other pieces of evidence might not discriminate among explanations at all. What matters is not the amount of evidence, but its contribution to adjudicating among alternative hypotheses. Further, even a single case may include many salient pieces of evidence. The noted methodologist Donald Campbell recognized the value of process-focused tools of inference when he abandoned his earlier criticism of
Andrew Bennett

case studies as lacking degrees of freedom, and argued in favor of a method similar to the process tracing under discussion here (Campbell 1975).

More concretely, process tracing involves several different kinds of empirical tests, focusing on evidence with different kinds of probative value. Van Evera (1997: 31–32) has distinguished four such tests that contribute in distinct ways to confirming and eliminating potential explanations. They are summarized briefly here, and will then be applied and illustrated throughout this chapter.

_Hoop tests_, which are central to the discussion below, can eliminate alternative hypotheses, but they do not provide direct supportive evidence for a hypothesis that is not eliminated. They provide a *necessary but not sufficient* criterion for accepting the explanation. The hypothesis must “jump through the hoop” just to remain under consideration, but success in passing a hoop test does not strongly affirm a hypothesis. Van Evera’s apt example of a hoop test is, “Was the accused in the state on the day of the murder?”

_Smoking gun tests_ strongly support a given hypothesis, but failure to pass such a test does not eliminate the explanation. They provide a *sufficient but not necessary* criterion for confirmation. As van Evera notes, a smoking gun

Table 10.1. Process Tracing: Four Tests for Causation

<table>
<thead>
<tr>
<th></th>
<th>Sufficient To Establish Causation&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Passing</strong></td>
<td>Passing affirms relevance of hypothesis but does not confirm it.</td>
<td>Passing confirms hypothesis.</td>
<td></td>
</tr>
<tr>
<td><strong>Failing</strong></td>
<td>Failing suggests hypothesis may not be relevant, but does not eliminate it.</td>
<td>Failing does not eliminate it.</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> The typology creates a new, two-dimensional framing of the alternative tests originally formulated by Van Evera (1997: 31–32).

<sup>b</sup> In this figure, “establishing causation,” as well as “confirming” or “eliminating” an hypothesis, obviously does not involve a _definitive_ test. Rather, as with any causal inference, qualitative or quantitative, it is a _plausible_ test in the framework of (a) this particular method of inference and (b) a specific data set.
in the suspect’s hands right after a murder strongly implicates the suspect, but the absence of such a gun does not exonerate a suspect.

Straw in the wind tests provide useful information that may favor or call into question a given hypothesis, but such tests are not decisive by themselves. They provide neither a necessary nor a sufficient criterion for establishing a hypothesis or, correspondingly, for rejecting it.

Finally, doubly decisive tests confirm one hypothesis and eliminate others. They provide a necessary and sufficient criterion for accepting a hypothesis. Just one doubly decisive piece of evidence may suffice, whereas many straw in the wind tests may still be indeterminate vis-à-vis alternative explanations. Van Evera’s example is a bank camera that catches the faces of robbers, thereby implicating those photographed and exonerating all others. He emphasizes that in the social sciences such tests are rare, yet a hoop test and a smoking gun test together accomplish the same analytic goal (1997: 32), a combination that is illustrated in the examples below.

In process tracing and in applying these tests, it is essential to cast the net widely in considering alternative explanations. Other standard injunctions advocate gathering diverse forms of data, being meticulous and even-handed in collecting and evaluating data, and anticipating and accounting for potential biases in the evidence (George and Bennett 2005, Bennett and Elman 2006). Further, as with all forms of causal inference, specific process tracing tests must be evaluated in relation to a wider body of evidence. These desiderata are especially important in process tracing on social and political phenomena for which participating actors have strong instrumental or ideational reasons for hiding or misrepresenting information about their behavior or motives.

Example: Why the Fashoda Crisis Did Not Result in War

Schultz provides excellent examples of the hoop test and smoking gun test in his analysis of the 1898 Fashoda crisis between Britain and France. This crisis arose over the confrontation between the two countries’ expeditionary forces as they raced to lay claim to the Upper Nile Valley. War was averted when France backed down. With the emergence of the inter-democratic peace research program in the last several decades, this episode has assumed special interest as a near war between two democracies, leading scholars to closely scrutinize explanations of its non-occurrence.

Schultz lays out three alternative explanations that scholars have offered for why the crisis was resolved without a war. Neorealists argue that France backed down simply because Britain’s military forces were far stronger, both in the region and globally (Layne 1994). Schultz rejects this explanation because it fails to survive a hoop test: it cannot explain why the crisis happened in the first place, why it lasted two months, and why it escalated
almost to the point of war, as it should have been obvious to France from the outset that Britain had military superiority (Schultz 2001: 177). A second argument, that democratic norms and institutions led to mutual restraint, also fails a hoop test in Schultz’s view. Whereas traditional democratic peace theorists emphasize the restraining power of democratic norms and institutions, the British public and British leaders were belligerent throughout the crisis in their rhetoric and actions toward France (Schultz 2001: 180–183).

Schultz then turns to his own explanation: democratic institutions force democratic leaders to reveal private information about their intentions, making it difficult for them to bluff in some circumstances but also making threats to use force more credible in others. In this view, democratic institutions reinforce the credibility of coercive threats when domestic opposition parties and publics support these threats, but they undermine the credibility of threats when domestic groups publicly oppose the use of force.

Schultz supports this explanation with smoking gun evidence. The credibility of Britain’s public commitment to take control of the region was resoundingly affirmed by the opposition Liberal Party leader Lord Rosebery (Schultz 2001: 188). Meanwhile, France’s Foreign Minister, Theophile Delcasse, initially voiced an intransigent position, but his credibility was quickly undermined by public evidence that other key French political actors were apathetic toward, or even opposed to, a war over Fashoda (Schultz 2001: 193). Within a matter of days after such costly signaling by both sides revealed Britain’s greater willingness and capability to fight for the Upper Nile, France began to back down, leading to a resolution of the crisis in Britain’s favor. In sum, the close timing of these events, following in the sequence predicted by Schultz’s theory, provides smoking gun evidence for his explanation; this, combined with the alternative explanations’ failures in hoop tests, makes Schultz’s explanation of the Fashoda case convincing.

Example: Expanding the Ends and Means of German Strategy in World War I

A second example shows how hoop tests and a smoking gun test help adjudicate among rival explanations for why Germany expanded both the ends and means of its wartime strategy in 1916-1917 even as it was becoming obvious that Germany was losing World War I. Goemans convincingly argues that four developments in 1916 made it increasingly evident to German leaders that they were unlikely to win the war: the German offensive at Verdun failed; Britain demonstrated its resolve—including its tolerance for casualties—in the battle of the Somme; Russia’s Brusilov offensive showed it could still fight; and Romania entered the war against Germany
(Goemans 2000: 89–93). Meanwhile, President Wilson’s diplomatic note to Germany in April 1916 after the sinking of the unarmed SS Sussex made it clear that the United States was almost certain to enter the war against Germany if German U-Boats sank any more merchant ships, which inhibited Germany from attacking merchantmen for the rest of the year.

Despite these developments, in late 1916 Germany escalated its terms for concluding the war, expanding its claims on Polish territory and increasing the territorial or diplomatic concessions it demanded from France, Belgium, and Russia (Goemans 2000: 98–106). Moreover, Germany returned to unrestricted submarine warfare in early 1917, even though the predictable consequence was that the United States, in quick response, entered the war.

Why did Germany expand the ends and means of its war strategy even as its probability of victory declined? Goemans evaluates five rival explanations. A first alternative—that Germany should have behaved as a unitary actor and responded only to international considerations—fails a *hoop test*, based on thorough evidence that Germany’s goals in the war expanded even though German leaders themselves understood that their prospects for victory had diminished. A second argument, that Germany was irrevocably committed to hegemony throughout the war, is also undercut by evidence that German war aims increased over time. Goemans rejects a third argument—Germany’s authoritarian government made it a “bad learner” impervious to evidence that it was losing the war—with ample indications that German leaders understood very well by late 1916 that their chances for victory were poor. A fourth explanation, that the change in Germany’s military leadership led to expanded military goals, begs the question of why Germany replaced its military leaders in the midst of the war (Goemans 2000: 74–75, 93–105).

Goemans then evaluates his own hypothesis: when semi-authoritarian governments, like that of Germany during World War I, believe they are losing a war, they are likely to respond with war strategies that preserve at least a small probability of resounding victory, even if such strategies have a high likelihood of abject defeat. Goemans argues that for leaders in such governments, the consequences of negotiating an end to a war on modestly concessionary terms are little different from those of losing the war outright. In either case, semi-authoritarian leaders are likely to lose their power and property (and perhaps even their lives) to domestic opponents who blame them for having demanded immense sacrifices from their societies in a losing cause. Thus, when evidence mounts that a semi-authoritarian state is losing in a war, its leaders have an incentive to gamble for resurrection and adopt riskier strategies that offer at least some slim hope of victory, even though they also increase the odds of utter defeat.

Goemans provides a *smoking gun test* for this argument in the case of Ger-
many’s escalating war aims. Among many other pieces of evidence, he quotes the German military leader Erich Ludendorff as arguing in a private letter that radical and unacceptable domestic political reforms would be required to stave off unrest if Germany were to negotiate a concessionary peace. Specifically, Ludendorff argued that the extension of equal voting rights in Prussia “would be worse than a lost war” (Goemans 2000: 114). This letter provides direct evidence of the German leadership’s desperation to avoid losing the war because of the political consequences for German leaders should they be blamed for having lost the war, and it thereby constitutes a smoking gun test that substantially validates Goemans’s main argument.

Example: The Peaceful End of the Cold War

The final example concerns use of the hoop, smoking gun, and straw in the wind tests to adjudicate among hypotheses about why the Soviet Union did not intervene militarily in the Eastern European revolutions of 1989. Three prominent accounts for the non-use of force, involving standard alternative explanatory perspectives in the international relations field, are: (1) a realist hypothesis, which emphasizes the changing material balance of power; (2) a domestic politics hypothesis, which focuses on the changing nature of the Soviet Union’s ruling coalition; and (3) an ideational hypothesis centered on Soviet leaders’ lessons from their recent experiences.

First, the most comprehensive realist/balance of power analysis of Soviet restraint in 1989 is offered by Brooks and Wohlforth (2000/2001; see also Wohlforth 1994/1995, Oye 1996). They argue that the decline in Soviet economic growth rates in the 1980s, combined with the Soviet Union’s high defense spending and its “imperial overstretch” in Afghanistan, led to Soviet foreign policy retrenchment in the late 1980s. Soviet leaders were constrained from using force in 1989 because this would have imposed large direct economic and military costs, risked economic sanctions from the West, and forced the Soviet Union to assume the economic burden of the large debts that Eastern European regimes had incurred to the West. In this view, changes in Soviet leaders’ ideas about foreign policy were largely determined by changes in their material capabilities.

Second, a domestic politics account has been well formulated by Snyder (1987/88). He argues that the long-term change in the Soviet economy from extensive development (focused on basic industrial goods) to intensive development (involving more sophisticated and information-intensive

---

4. I use this example in part because it involves my own research, making it easier to reconstruct the steps involved in the process tracing. See Bennett (1999, 2003, 2005).
goods and services) shifted the ruling Soviet coalition from a military/heavy-industry/party complex to a power bloc centered in light industry and the intelligentsia. This led the Soviet Union to favor improved ties to the West to gain access to technology and trade, and any Soviet use of force in Eastern Europe in 1989 would have damaged Soviet economic relations with the West.

The third line of argument maintains that Soviet leaders learned lessons from their unsuccessful military interventions in Afghanistan and elsewhere that led them to doubt the efficacy of using force to try to resolve political problems like the Eastern Europeans’ demands for independence from the Soviet Union in 1989. The Soviet Union invaded Afghanistan in December 1979 and kept between 80,000 and 100,000 troops there for a decade, with over 14,000 Soviet soldiers killed and 53,000 injured. When even this effort and substantial economic aid failed to make the communist party of Afghanistan capable of defending itself, Soviet leaders withdrew their military forces in February 1989. The learning explanation argues that this experience made Soviet leaders unwilling to use force nine months later to keep in power Eastern European leaders who by that time faced strong public opposition.

While scholars agree that the variables highlighted by all of these hypotheses contributed to the non-use of force in 1989, there remains considerable disagreement on how these variables interacted and their relative causal weight. Brooks and Wohlforth, for example, disagree with the “standard view” that “even though decline did prompt change in Soviet foreign policy, the resulting shift could just as easily have been toward aggression or a new version of muddling through . . . and that other factors played a key role in resolving this uncertainty” (2002: 94). In contrast, I assert that this standard interpretation is persuasive and maintain that were it not for other factors, the economic decline of the Soviet Union relative to the West could indeed have led to renewed Soviet aggression or to more years of muddling through. Specifically, I argue that although changes in the material balance of power made Soviet leaders more open to new ideas, the particular lessons Soviet leaders drew from their uses of force in the 1970s and 1980s greatly influenced the timing and direction of changes in Soviet foreign policy.

What kinds of evidence can adjudicate among these hypotheses? In introducing a symposium on competing views on these hypotheses, Tannenwald (2005) poses three questions for judging them: (1) Did ideas correlate with the needs of the Soviet State, actors’ personal material interests, or actors’ personal experiences and the information to which they were

exposed? (2) Did material change precede or follow ideational change? (3) Do material or ideational factors better explain which ideas won out? Each of these questions creates opportunities for process tracing tests.

Focusing on the first question, about the correlation of policy positions with material versus ideational variables, we find some evidence in favor of each explanation. Citing Soviet Defense Minister Yazov and others, Brooks and Wohlforth argue that Soviet conservatives and military leaders did not question Gorbachev’s concessionary foreign policies because they understood that the Soviet Union was in dire economic straits and needed to reach out to the West. They also point to ample evidence that Gorbachev argued that Soviet economic decline created a need for better relations with the West (Brooks and Wohlforth 2000/2001). Their explanation thus satisfies a *hoop test*: given the salience of both economic issues and relations with the West, Brooks and Wohlforth’s argument would be unsustainable without considerable evidence that Soviet leaders linked the two in their public and private statements.

However, Robert English suggests that the evidence we have employed in this *hoop test* is not definitive, and he points to other statements by Soviet conservatives indicating opposition to Gorbachev’s foreign policies. He concludes that ‘whatever one believes about the old thinkers’ acquiescence in Gorbachev’s initiatives, it remains inconceivable that they would have launched similar initiatives without him’ (English 2002: 78). In this view, much of the evidence linking material decline to Soviet retrenchment depends on Gorbachev’s individual views and the political institutions that gave him power, rather than any direct and determinative tie between material decline and specific foreign policies.

Two other *hoop tests* yield more definitive evidence against Snyder’s sectoral interest group hypothesis and in favor of the learning hypothesis. Consistent with Snyder’s argument, Soviet military leaders at times argued against defense spending cuts, and the conservatives who attempted a coup against Gorbachev in 1990 represented the Stalinist coalition of the military and heavy industry. Soviet Conservatives, however, did not argue that force should have been used to prevent the dissolution of the Warsaw Pact in 1989, even after they had fallen from power in 1990 and had little to lose (Bennett 2005: 104). Indeed, military leaders were among the early skeptics regarding the use of force in Afghanistan, and many prominent officers with personal experience in Afghanistan resigned their commissions rather than participating in the 1994–1997 Russian intervention in Chechnya (Bennett 1999: 339–340). This suggests that the learning explanation has survived a difficult *hoop test* by correctly anticipating that those military officers who personally experienced failure in Afghanistan would be among the opponents rather than the supporters of using force in later circumstances.
Concerning Tannenwald's second question, about the timing of material and ideational change, Brooks and Wohlforth have not indicated precisely the time frame within which material decline would have allowed or compelled Soviet foreign policy change, stating only that material incentives shape actions over the "longer run" (2002: 97). This suggests that the timing of changes in Soviet policy in relation to that of changes in the material balance of power is at best a straw in the wind test. Brooks and Wohlforth's logic allows for the possibility that the Soviet Union could profitably have let go of its Eastern European empire in 1973. By that time, nuclear parity guaranteed the Soviet Union's security from external attack, and high energy prices meant that the Soviet Union could have earned more for its oil and natural gas from world markets than from Eastern Europe. Moreover, the sharpest decline in the Soviet economy came after 1987, by which time Gorbachev had already begun to signal to governments in Eastern Europe that he would not use force to rescue them from popular opposition (Brown 1996: 249). The timing of changes in Soviet policy therefore does not lend strong support for the "material decline" hypothesis.

The timing suggested by the ideational explanation coincides much more closely with actual changes in Soviet foreign policy. Despite slow Soviet economic growth, Soviet leaders were optimistic about the use of force in the developing world in the late 1970s due to the ease with which they inflicted a costly defeat on the United States in Vietnam, but they became far more pessimistic regarding the efficacy of force as their failure in Afghanistan deepened through the 1980s (Bennett 1999). Furthermore, changes in Soviet leaders' public statements generally preceded changes in Soviet foreign policy, suggesting that the driving factor was ideational change, rather than material interests justified by ad hoc and post hoc changes in stated ideas. In this regard, the ideational explanation survives a hoop test: if changes in Soviet leaders' ideas motivated changes in their policies, rather than being merely rationalizations for policy changes adopted for instrumental reasons, then changes in these ideas had to precede those in behavior (Bennett 1999: 351–2).

Tannenwald's third question, on why some ideas won out over others, is the one most effectively addressed by hoop tests. Here, although Snyder does not specifically apply his domestic politics argument to Soviet restraint in the use of force in 1989, his contention that the material interests of different sectors were the driving factor in Soviet policy appears to fail a hoop test (Snyder 1990). Outlining in early 1988 the (then) hypothetical future events that could in his view have caused a resurgence of the Stalinist coalition of the military and heavy industry, Snyder argued that the rise of anti-reform Soviet leaders would become much more likely if Gorbachev's reforms were discredited by poor economic performance and if the Soviet Union faced "a hostile international environment in which SDI [the Strate-
While both material and ideational considerations played a role, there is reason to believe that at least in one respect the former was not a factor in Gorbachev’s thinking in the fall of 1989. In a meeting on October 31, 1989, just ten days before the Berlin Wall fell, Gorbachev was reportedly “astonished” at hearing from East German leader Egon Krenz that East Germany owed the West $26.5 billion, almost half of which had been borrowed in 1989 (Zelikow and Rice 1995: 87). Thus, while Gorbachev was certainly concerned about Soviet economic performance, the claim that he was in part inhibited from using force in Eastern Europe because of the region’s external debts appears to have failed a hoop test because almost up until the Berlin Wall fell, Gorbachev did not even know the extent of these debts.

In sum, the material decline explanation passes a hoop test by showing that a wide range of Soviet leaders acknowledged Soviet decline, and a straw
in the wind test on the timing of changes in Soviet foreign policy, but the variant of this explanation that stresses East German debts as a factor preventing the Soviet use of force in 1989 fails a hoop test. The learning explanation survives hoop tests in its expectations on which actors would espouse which foreign policy views, on the timing of changes in Soviet ideas and policies, and on why some ideas prevailed over others. The sectoral domestic politics explanation emerges as the weakest, having failed hoop tests on its predicted correlation of policy views and material interests and its expectations on which ideas would win out in which contexts.

**CONCLUSION**

Through process tracing, scholars can make valuable inferences if they have the right kind of evidence. “Right kind” means that some types of evidence have far more probative value than others. The evidence must strongly discriminate between alternative hypotheses in the ways discussed above. The idea of hoop tests, smoking gun tests, doubly decisive tests, and straw in the wind tests brings into focus some of the key ways in which this discrimination occurs. What matters is the relationship between the evidence and the hypotheses, not the number of pieces of evidence.

Process tracing is not a panacea for causal inference, as all methods of causal inference are potentially fallible. Researchers could fail to include an important causal variable in their analyses. Available evidence may not discriminate strongly between competing and incompatible explanations. Actors may go to great lengths to obscure their actions and motivations when these are politically sensitive, biasing available evidence. Yet with appropriate evidence, process tracing is a powerful means of discriminating among rival explanations of historical cases even when these explanations involve numerous variables.
One type of scientific inquiry involves the analysis of large data sets, often using statistical models and formal tests of hypotheses. A moment’s thought, however, shows that there must be other types of scientific inquiry. For instance, something has to be done to answer questions like the following. How should a study be designed? What sorts of data should be collected? What kind of a model is needed? Which hypotheses should be formulated in terms of the model and then tested against the data?

The answers to these questions frequently turn on observations, qualitative or quantitative, that give crucial insights into the causal processes of interest. Such observations generate a line of scientific inquiry, or markedly shift the direction of the inquiry by overturning prior hypotheses, or provide striking evidence to confirm hypotheses. They may well stand on their own rather than being subsumed under the systematic data collection and modeling activities mentioned above.

Such observations have come to be called “Causal Process Observations” (CPOs). These are contrasted with the “Data Set Observations” (DSOs) that are grist for statistical modeling (Collier, Brady, and Seawright, chap. 9, this volume). My object in this essay is to illustrate the role played by CPOs, and qualitative reasoning more generally, in a series of well-known episodes drawn from the history of medicine.

Why is the history of medicine relevant to us today? For one thing, medical researchers frequently confront observational data that present familiar challenges to causal inference. For another, distance lends perspective,
allowing gains and losses to be more sharply delineated. The examples show that an impressive degree of rigor can be obtained by combining qualitative reasoning, quantitative analysis, and experiments when those are feasible. The examples also show that great work can be done by spotting anomalies, and trying to understand them.

QUALITATIVE REASONING: CASE STUDIES FROM EPIDEMIOLOGY

Jenner and Vaccination

The setting is the English countryside in the 1790s. Cowpox, as will be clear from the name, is a disease of cows. The symptoms include sores on the teats. Those who milk the cows often became infected, with sores on their hands; by the standards of the time, the illness is rarely serious. In contrast, smallpox is one of the great killers of the 18th century.

In 1796, Edward Jenner took some matter from a cowpox sore on the hand of dairymaid Sarah Nelmes, and inserted it into the arm of an eight-year-old boy, “by means of two superficial incisions, barely penetrating the cutis, each about half an inch long.” The boy was “perceptibly indisposed” on the ninth day, but recovered the following day. Six weeks later, Jenner inoculated him with matter taken from a smallpox pustule, “but no disease followed” (Jenner 1798, Case XVII).

Jenner published 23 case studies to demonstrate the safety and efficacy of “vaccination,” as his procedure came to be called: vacca is the Latin term for cow, and vaccinia is another term for cowpox. Despite initial opposition, vaccination became standard practice within a few years, and Jenner achieved international fame. By 1978, smallpox had been eradicated.

What led Jenner to try his experiment? The 18th century view of disease was quite different from ours. The great Scottish doctor of the time, William Cullen, taught that most diseases were “caused by external influences—climate, foodstuffs, effluvia, humidity, and so on—and . . . the same external factors could cause different diseases in different individuals, depending on the state of the nervous system” (Porter 1997, 262).

Despite such misconceptions, it was known that smallpox could somehow be communicated from one person to another; moreover a person who contracted smallpox and survived was generally immune to the disease from that point on. As a preventive measure, patients could be deliberately infected (through scratches on the skin) with minute quantities of material taken from smallpox pustules, the idea being to induce a mild case of the disease that would confer immunity later.

This procedure was called “inoculation” or “variolation.” It was not free of risk: serious disease was sometimes caused in the patient, and in people
who came into contact with the patient (smallpox is highly contagious). On the other hand, failure to inoculate could easily lead to death from smallpox.

By the early part of the 18th century, variolation had reached England. Jenner was a country doctor who performed variolations. He paid attention to two crucial facts—although these facts were not explicable in terms of the medical knowledge of his time. (i) People who had the cowpox never seemed to contract smallpox afterwards, whether they had been inoculated or not. (ii) Some of his patients who had been ill with cowpox in the past still wanted to be inoculated; such patients reacted very little to inoculation—

What renders the Cox-pox virus so extremely singular, is, that the person who has been thus affected is for ever after secure from the infection of the Small Pox; neither exposure to the variolous effluvia, nor the insertion of the matter into the skin, producing this distemper. (Jenner 1798, 6)

These two facts led him to a hypothesis: cowpox created immunity against smallpox. That is the hypothesis he tested, observationally and experimentally, as described above. In our terminology, Jenner vaccinated a boy (Case XVII) who showed no response to subsequent inoculation. Immunity to smallpox had been induced by the vaccination.

By "virus,” Jenner probably meant “contagious matter,” that being a standard usage in his time. Viruses in the modern sense were not to be discovered for another century. By a curious twist, smallpox and cowpox are viral diseases in our sense too.

**Semmelweis and Puerperal Fever**

The time is 1844 and the place is Vienna. The discovery of microbes as the cause of infectious disease will not be made for some decades. Ignac Semmelweis is an obstetrician in the First Division of the Lying-in Hospital, where medical students are trained. (Midwives are trained in the Second Division.) Pregnant women are admitted to one division or the other, according to the day of the week that they come to the hospital, in strict alternation. Mortality from “puerperal fever” is much higher in the First Division (Semmelweis 1981 [1860]: 356).

Eventually, Semmelweis discovers the cause. The medical students are doing autopsies, and then examining the “puerperae” (women who are giving birth, or who have just given birth). "Cadaveric particles" are thus transferred to the women, entering the bloodstream and causing infection. In 1847, Semmelweis institutes the practice of disinfection, and mortality plummets (Semmelweis 1981 [1860]: 393–4).
But how did Semmelweis make his discovery? To begin with, he had to reject conventional explanations, including “epidemic influences,” which meant something different then:

Epidemic influences . . . are to be understood [as] certain hitherto inexplicable, atmospheric, cosmic, telluric changes, which sometimes disseminate themselves over whole countriesides, and produce childbed fever in individuals predisposed thereto by the puerperal state. [“Telluric” means earthly.] Now, if the atmospheric-cosmic-telluric conditions of the City of Vienna are so disposed that they cause puerperal fever in individuals susceptible thereto as puerperae, how does it happen that these atmospheric-cosmic-telluric conditions over such a long period of years have carried off individuals disposed thereto as puerperae in the First Clinic, while they have so strikingly spared others also in Vienna, even in the same building in the Second Division and similarly vulnerable as puerperae?” (Semmelweis 1981 [1860]: 357).

The reasoning is qualitative; and similar qualitative arguments dispose of other theories—diet, ventilation, use of hospital linens, and so forth.

Now he has to discover the real cause. In 1847, his revered colleague Professor Kolletschka is accidentally cut with a knife used in a medico-legal autopsy. Kolletschka becomes ill, with symptoms remarkably similar to puerperal fever; then he dies. Again, qualitative analysis is crucial. Close attention to symptoms and their progression is used to identify Kolletschka’s illness with puerperal fever (Semmelweis 1981 [1860]: 391). Tracing of causal processes comes into play as well:

Day and night this picture of Kolletschka’s disease pursued me. . . . I was obliged to acknowledge the identity of the disease, from which Kolletschka died, with that disease of which I saw so many puerperae die. . . . I must acknowledge, if Kolletschka’s disease and the disease from which I saw so many puerperae die, are identical, then in the puerperae it must be produced by the self-same engendering cause, which produced it in Kolletschka. In Kolletschka, the specific agent was cadaveric particles, which were introduced into his vascular system [the bloodstream]. I must ask myself the question: Did the cadaveric particles make their way into the vascular systems of the individuals, whom I had seen die of an identical disease? This question I answer in the affirmative. (Semmelweis 1981 [1860]: 391–2)

The source of the infectious agent could also be a wound in a living person (Semmelweis 1981 [1860]: 396). Once the cause is discovered, the remedy is not far away: eliminate the infectious particles from the hands that will examine the puerperae. Washing with soap and water is insufficient, but disinfection with chlorine compounds is sufficient (Semmelweis 1981 [1860]: 392–96).

Semmelweis’ work was accepted by few of his contemporaries, due in
part to his troubled and disputatious personality, although his picture of the disease was essentially correct. Puerperal fever is a generalized infection, typically caused by bacteria in the group *Streptococcus pyogenes*. These bacteria enter the blood-stream through wounds suffered during childbirth (for instance, at the site where the placenta was attached). Puerperal fever can be—and today it generally is—avoided by proper hygiene.

**Snow and Cholera**

John Snow was a physician in Victorian London. In 1854, he demonstrated that cholera was an infectious disease, which could be prevented by cleaning up the water supply. The demonstration took advantage of a natural experiment. A large area of London was served by two water companies. The Southwark and Vauxhall company distributed contaminated water, and house-holds served by it had a death rate “between eight and nine times as great as in the houses supplied by the Lambeth company,” which supplied relatively pure water (Snow 1965 [1855]: 86, data in Table IX).

What led Snow to design the study and undertake the arduous task of data collection? To begin with, he had to reject the explanations of cholera epidemics that were conventional in his time. The predominant theory attributed cholera to “miasmas,” that is, noxious odors—especially odors generated by decaying organic material. Snow makes qualitative arguments against such explanations:

> [Cholera] travels along the great tracks of human intercourse, never going faster than people travel, and generally much more slowly. In extending to a fresh island or continent, it always appears first at a sea-port. It never attacks the crews of ships going from a country free from cholera, to one where the disease is prevailing, till they have entered a port, or had intercourse with the shore. Its exact progress from town to town cannot always be traced; but it has never appeared except where there has been ample opportunity for it to be conveyed by human intercourse. (Snow 1965 [1855]: 2)

These phenomena are easily understood if cholera is an infectious disease, but hard to explain on the miasma theory. Similarly,

The first case of decided Asiatic cholera in London, in the autumn of 1848, was that of a seaman named John Harnold, who had newly arrived by the *Elbe* steamer from Hamburgh, where the disease was prevailing. . . . Now the next case of cholera, in London, occurred in the very room in which the above patient died. (Snow 1965 [1855]: 3)

The first case was infected in Hamburgh; the second case was infected by contact with dejecta from the first case, on the bedding or other furnishings.
in that fatal room. The miasma theory, on the other hand, does not provide
good explanations.

Careful observation of the disease led to the conclusion “that cholera
invariably commences with the affection of the alimentary canal” (Snow
1965, 10). A living organism enters the body, as a contaminant of water or
food, multiplies in the body, and creates the symptoms of the disease.
Many copies of the organism are expelled from the body with the dejecta,
contaminate water or food, then infect other victims. The task is now to
prove this hypothesis.

According to Sir Benjamin Ward Richardson, who wrote the introduction
to Snow’s book, the decisive proof came during the Broad Street epidemic
of 1854:

[Snow] had fixed his attention on the Broad Street pump as the source and
centre of the calamity. He advised the removal of the pump-handle as the
grand prescription. The vestry [in charge of the pump] was incredulous, but
had the good sense to carry out the advice. The pump-handle was removed
and the plague was stayed. (Snow 1965 [1855]: xxxvi)

The pump-handle as the decisive test is a wonderful fable, which has
beguiled many a commentator.

What are the facts? Contamination at the pump did cause the epidemic,
Snow recommended closing the pump, his advice was followed, and the
epidemic stopped. However, the epidemic was stopping anyway. Closing
the pump had no discernible effect: the episode proves little. Snow explains
this with great clarity (Snow 1965 [1855]: 40–55, see esp. Table I on p. 49
and the conclusory paragraph on pp. 51–2). Richardson’s account is there-
fore a classic instance of post hoc, ergo propter hoc.

The reality is more interesting than the fable. Snow was intimately famil-
lar with the Broad Street area, because of his medical practice. He says,

As soon as I became acquainted with the situation and extent of this irruption
of cholera, I suspected some contamination of the water of the much-fre-
quented street-pump in Broad Street. . . . but on examining the water, on the
evening of 3rd September, I found so little impurity in it of an organic nature,
that I hesitated to come to a conclusion. (Snow 1965 [1855]: 38–39)

Snow had access to the death certificates at the General Register Office,
and drew up a list of the cholera fatalities registered shortly before his
inspection of the pump. He then made a house-to-house canvass (the death
certificate shows the address of the deceased), and discovered that the cases
clustered around the pump, confirming his suspicion. Later, he made a
more complete tally of cholera deaths in the area. His “spot map” displays
the locations of cholera fatalities during the epidemic, and the clustering is
On Types of Scientific Inquiry

apparent from the map (Snow 1965 [1855]: 44–45; Cholera Inquiry Committee 1855: 106–9).

However, there were a number of exceptions that had to be explained. For example, there was a brewery near the pump; none of the workers contracted the disease: why not? First, the workers drank beer; second, if water was desired, there was a pump on the premises (Snow 1965 [1855]: 10). For another example, a lady in Hampstead contracted cholera: why? As it turned out, she liked the taste of the water from the Broad Street pump, and had it brought to her house (Snow 1965 [1855]: 44). Snow gives many other such examples.

Snow’s work on the Broad Street epidemic illustrates the power of case studies. His refutation of the usual explanations for cholera, and the development of his own explanation, are other indicators of the power of qualitative reasoning. The analysis of his natural experiment, referred to above, shows the power of simple quantitative methods and good research design. This was the great quantitative test of his theory that cholera was a waterborne infectious disease.

In designing the quantitative study, however, Snow made some key qualitative steps: (i) seeing that conventional theories were wrong, (ii) formulating the water hypothesis, and (iii) noticing that in 1852, the Lambeth company moved its intake pipe to obtain relatively pure water, while Southwark and Vauxhall continued to draw heavily contaminated water. It took real insight to see—a priori rather than a posteriori—that this difference between the companies allowed the crucial study to be done.

Snow’s ideas gained some circulation, especially in England. However, widespread acceptance was achieved only when Robert Koch isolated the causal agent (*Vibrio cholerae*, the comma-shaped bacillus) during the Indian epidemic of 1883. Even then, there were dissenters, with catastrophic results in the Hamburg epidemic of 1892 (Evans 1987).

Inspired by Koch and Louis Pasteur, there was a great burst of activity in microbiology during the 1870s and 1880s. The idea that microscopic life forms could arise by spontaneous generation was cast aside, and the germ theory of disease was given solid experimental proof. Besides the cholera vibrio, the bacteria responsible for anthrax (*Bacillus anthracis*) and for tuberculosis (*Mycobacterium tuberculosis*) were isolated, and a vaccine was developed against rabies. However, as we shall see in a moment, these triumphs made it harder to solve the riddle of beriberi. Beriberi is a deficiency disease, but the prestige of the new microbiology made investigators suspicious of any explanation that did not involve microorganisms.

Eijkman and Beriberi

Beriberi was endemic in Asia, from about 1750 until 1930 or so. Today, the cause is known. People need minute amounts (about one part per mil-
lion in the diet) of a vitamin called “thiamin.” Many Asians eat a diet based on rice, and white rice is preferred to brown.

Thiamin in rice is concentrated in the bran—the skin that gives rice its color. White rice is obtained by polishing away the skin, and with it most of the thiamin; what is left is further degraded by cooking. The diet is then deficient in thiamin, unless supplemented by other foods rich in that substance. Beriberi is the sequel.

In 1888, knowledge about vitamins and deficiency diseases lay decades in the future. That year, Christiaan Eijkman—after studying microbiology with Koch in Berlin—was appointed director of the Dutch Laboratory for Bacteriology and Pathology in the colony of Java, near the city now called Jakarta. His research plan was to show that beriberi was an infectious disease, with Koch’s methods for the proof.

Eijkman tried to infect rabbits and then monkeys with blood drawn from beriberi patients. This was unsuccessful. He then turned to chickens. He tried to infect some of the birds, leaving others as controls. After a time, many of his chickens came down with polyneuritis, which he judged to be very similar to beriberi in humans. (“Polyneuritis” means inflammation of multiple nerves.)

However, the treated chickens and the controls were equally affected. Perhaps the infection spread from the treated chickens to the controls? To minimize cross infection, he housed the treated chickens and the controls separately. That had no effect. Perhaps his whole establishment had become infected? To eliminate this possibility, he started work on another, remote experimental station—at which point, the chickens began recovering from the disease.

[Eijkman] wrote “something struck us that had escaped our attention so far.” The chickens had been fed a different diet during the five months in which the disease had been developing. In that period (July through November 1889), the man in charge of the chickens had persuaded the cook at the military hospital, without Eijkman being aware of it, to provide him with leftover cooked [white] rice from the previous day, for feeding to the birds. A new cook, who started duty on 21 November, had refused to continue the practice. Thirty years later, Eijkman was to say that “[the new cook] had seen no reason to give military rice to civilian hens.” (Carpenter 2000, 38)

In short, the chickens became ill when fed cooked, polished rice; they recovered when fed uncooked, unpolished rice. This was an accidental experiment, arranged by the cooks. One of Eijkman’s great insights was paying attention to the results, because the cooks’ experiment eventually changed the understanding of beriberi.

Eijkman’s colleague Adolphe Vorderman undertook an observational study of prisons, to confirm the relevance to humans. Where prisoners were
fed polished rice, beriberi was common; with a diet of unpolished rice, beriberi was uncommon. Beriberi is a deficiency disease, not an infectious disease.

The evidence may seem compelling, but that is because we know the answer. At the time, the picture was far from clear. Eijkman himself thought that white rice was poisonous, the bran containing the antidote. Later, he was to reverse himself: beriberi is an infectious disease, although a poor diet makes people (and chickens) more vulnerable to infection.

In 1896, Gerrit Grijns took over Eijkman’s lab (Eijkman suffered from malaria, and had to return to Holland). Among other contributions, after a long series of careful experiments, Grijns concluded that beriberi was a deficiency disease, the missing element in the diet being concentrated in rice bran—and in other foods like mung beans.

In 1901, Grijn’s colleague Hulshoff Pol ran a controlled experiment at a mental hospital, showing that mung beans prevented or cured beriberi. In three pavilions out of twelve, the patients were fed mung beans; in three pavilions, other green vegetables. In three pavilions, there was intensive disinfection, and three pavilions were used as controls. The incidence of beriberi was dramatically lower in the pavilions with mung beans.

Still, medical opinion remained divided. Some public health professionals accepted the deficiency hypothesis. Others continued to favor the germ theory, and still others thought the cause was an inanimate poison. It took another ten years or so to reach consensus that beriberi was a deficiency disease, which could be prevented by eating unpolished rice, or enriching the diet in other ways. From a public health perspective, the problem of beriberi might be solved, but the research effort turned to extracting the critical active ingredient in rice bran—no mean challenge, since there is about one teaspoon of thiamin in a ton of bran.

Around 1912, Casimir Funk coined the term “vitamines,” later contracted to vitamins, as shorthand for “vital amines.” The claim that he succeeded in purifying thiamin may be questionable. But he did guess that beriberi and pellagra were deficiency diseases, which could be prevented by supplying trace amounts of organic nutrients.

By 1926, B. C. P. Jansen and W. F. Donath had succeeded in extracting thiamin (vitamin B1) in pure crystal form. Ten years later, Robert R. Williams and his associates managed to synthesize the compound in the lab. In the 1930s, there were still beriberi cases in the East—and these could be cured by injecting a few milligrams of the new vitamin B1.

**Goldberger and Pellagra**

Pellagra was first observed in Europe in the eighteenth century by a Spanish physician, Gaspar Casal, who found that it was an important cause of ill-
health, disability, and premature death among the very poor inhabitants of the Asturias. In the ensuing years, numerous . . . authors described the same condition in northern Italian peasants, particularly those from the plain of Lombardy. By the beginning of the nineteenth century, pellagra had spread across Europe, like a belt, causing the progressive physical and mental deterioration of thousands of people in southwestern France, in Austria, in Rumania, and in the domains of the Turkish Empire. Outside Europe, pellagra was recognized in Egypt and South Africa, and by the first decade of the twentieth century it was rampant in the United States, especially in the south. . . . (Roe 1973: 1)

Pellagra seemed to hit some villages much harder than others. Even within affected villages, many households were spared, but some had pellagra cases year after year. Sanitary conditions in diseased households were primitive: flies were everywhere. One blood-sucking fly (*Simulium*) had the same geographical range as pellagra, at least in Europe; and the fly was most active in the spring, just when most pellagra cases developed. Many epidemiologists concluded the disease was infectious, and—like malaria or yellow fever—was transmitted from one person to another by insects.

Joseph Goldberger was an epidemiologist working for the U. S. Public Health Service. In 1914, he was assigned to work on pellagra. Despite the climate of opinion described above, he designed a series of observational studies and experiments showing that pellagra was caused by a bad diet, and is not infectious. The disease could be prevented or cured by foods rich in what Goldberger called the P-P (pellagra-preventive) factor.

By 1926, he and his associates had tentatively identified the P-P factor as part of the vitamin B complex. By 1937, C. A. Elvehjem and his associates had identified the P-P factor as niacin, also called vitamin B3 (this compound had been discovered by C. Huber around 1870, but its significance had not been recognized). Since 1940, most of the flour sold in the United States has been enriched with niacin, among other vitamins.

Niacin occurs naturally in meat, milk, eggs, some vegetables, and certain grains. Corn, however, contains relatively little niacin. In the pellagra areas, the poor ate corn—and not much else. Some villages and some households were poorer than others, and had even more restricted diets. That is why they were harder hit by the disease. The flies were a marker of poverty, not a cause of pellagra.

What prompted Goldberger to think that pellagra was a deficiency disease rather than an infectious disease? In hospitals and asylums, the inmates frequently developed pellagra, the attendants almost never—which is unlikely if the disease is infectious, because the inmates could infect the attendants. This observation, although far from definitive, set Goldberger on the path to discovering the cause of pellagra and methods for prevention or cure. The qualitative thinking precedes the quantitative investigation.
Pellagra is virtually unknown in the developed world today, although it remains prevalent in some particularly poor countries.

**Fleming and Penicillin**

Alexander Fleming was working at St. Mary's Hospital in London, under the direction of Sir Almroth Wright, studying the life cycle of staphylococcus (bacteria that grow in clusters, looking under the microscope like clusters of grapes). Fleming had a number of plates on which he was growing staphylococcus colonies. He left the plates in a corner of his office for some weeks while he was on holiday. When he returned, one of the plates had been contaminated by mold. So far, this is unremarkable. He noticed, however, "that around a large colony of a contaminating mould the staphylococcus colonies became transparent and were obviously undergoing lysis" (Fleming 1929: 226).

Bacteria "lyse" when their cell walls collapse. What caused the lysis? Rather than discarding the plate—the normal thing to do—Fleming thought that the lysis was worth investigating. He did so by growing the mold in broth, watching its behavior, and trying filtered broth on various kinds of bacteria. The mold, a species of *Penicillium*, generated a substance that "to avoid the repetition of the rather cumbersome phrase ‘mould broth filtrate’ [will be named] ‘penicillin’ " (Fleming 1929: 227). It was the penicillin that caused the bacteria to lyse. Fleming showed that penicillin destroyed—or at least inhibited the growth of—many kinds of bacteria besides staphylococcus.

Penicillin’s therapeutic potential went unrealized until Howard Florey and his associates at Oxford took up the research in 1938 and found processes for purification and larger-scale production. Due to the exigencies of World War II, much of the work was done in the U. S., where a strain of *Penicillium* that gave high yields was found on a moldy cantaloupe at a market in Peoria. (Industrial-scale development was being done at a nearby Department of Agriculture laboratory under the direction of Kenneth Raper, and people were encouraged to bring in moldy fruit for analysis.)

Penicillin was widely used to treat battlefield injuries, largely preventing gangrene, for example. Along with the sulfa drugs (prontosil was discovered by Gerhard Domagk in 1932) and streptomycin (discovered by Selman Waksman in 1944), penicillin was among the first of the modern antibiotics.

**CONCLUSIONS**

In the health sciences, there have been enormous gains since the time of Jenner, many of which are due to statistics. Snow's analysis of his natural
experiment shows the power of quantitative methods and good research design. Semmelweis’ argument depends on statistics; so too with Goldberger. On the other hand, as the examples demonstrate, substantial progress also derives from informal reasoning and qualitative insights. Recognizing anomalies is important; so is the ability to capitalize on accidents. Progress depends on refuting conventional ideas if they are wrong, developing new ideas that are better, and testing the new ideas as well as the old ones. The examples show that qualitative methods can play a key role in all three tasks.

In Fleming’s lab, chance circumstances generated an anomalous observation. Fleming resolved the anomaly and discovered penicillin. Semmelweis used qualitative reasoning to reject older theories about the cause of puerperal fever, to develop a new theory from observations on a tragic accident, and to design an intervention that would prevent the disease. The other examples lead to similar conclusions.

What are the lessons for methodologists in the 21st century? Causal inference from observational data presents many difficulties, especially when underlying mechanisms are poorly understood. There is a natural desire to substitute intellectual capital for labor, and an equally-natural preference for system and rigor over methods that seem more haphazard. These are possible explanations for the current popularity of statistical models.

Indeed, far-reaching claims have been made for the superiority of a quantitative template that depends on modeling—by those who manage to ignore the far-reaching assumptions behind the models. However, the assumptions often turn out to be unsupported by the data (Duncan 1984a; Berk 2004; Freedman 2005; chaps. 1 and 9, this volume). If so, the rigor of advanced quantitative methods is a matter of appearance rather than substance.

The historical examples therefore have another important lesson to teach us. Scientific inquiry is a long and tortuous process, with many false starts and blind alleys. Combining qualitative insights and quantitative analysis—and a healthy dose of skepticism—may provide the most secure results.

FURTHER READING

Brady, Collier, and Seawright (chaps. 1 and 9, this volume) compare qualitative and quantitative methods for causal inference in the social sciences. As they point out,

it is difficult to make causal inferences from observational data, especially when research focuses on complex political processes. Behind the apparent
On Types of Scientific Inquiry

On Types of Scientific Inquiry

precision of quantitative findings lie many potential problems concerning equivalence of cases, conceptualization and measurement, assumptions about the data, and choices about model specification. (22 this volume)

These authors recommend using a diverse mix of qualitative and quantitative techniques in order to exploit the available information; no particular set of tools is universally best. Causal process observations (including anomalies and results of accidental experiments, even experiments with \( N = 1 \)) can be extremely helpful, as they were in the epidemiological examples discussed here.

The role of anomalies in political science is also discussed by Rogowski (chap. 5, this volume). He suggests that scholars in that field may be excessively concerned with hypothesis testing based on statistical models. Scholars may underestimate the degree to which the discovery of anomalies can overturn prior hypotheses and open new avenues of investigation. Anomalies that matter have been discovered in case studies—even when the cases have been selected in ways that do considerable violence to large-\( N \) canons for case selection. He also suggests that failure to search for anomalies can lead to a kind of sterility in research programs.

Scientific progress often begins with inspired guesswork. On the other hand, if guesses cannot be verified, progress may be illusory. For example, Snow (1965 [1855]: 125–33) theorized that—by analogy with cholera—plague, yellow fever, dysentery, typhoid fever, and malaria (which he calls “ague” or “intermittent fever”) were waterborne infectious diseases. His supporting arguments were thin. As it turns out, these diseases are infectious; however, only dysentery and typhoid fever are waterborne.

Proof for dysentery and typhoid fever, and disproof for the other diseases, was not to come in Snow’s lifetime. Although William Budd (1873) made a strong case on typhoid fever, reputable authors of the late 19th century still denied that such diseases were infectious (Bristowe and Hutchinson 1876: 211, 629; Bristowe et al. 1879: 102–3). In the following decades, evidence from epidemiology and microbiology settled the issue.

Plague is mainly spread by fleas, although transmission by coughing is also possible. The causal agent is the bacterium *Yersinia pestis*. Yellow fever and malaria are spread by mosquitoes. Yellow fever is caused by a virus. Malaria is caused by several species of *Plasmodium*, one-celled organisms with a nucleus and an extravagantly complicated life-cycle spent partly in humans and partly in mosquitoes. The medieval Black Death is usually identified with modern plague, but this is still contested by some scholars (Nutton 2008).

Buck et al. (1989) reprints many of the classic papers in epidemiology; some classic errors are included too. Porter (1997) is a standard reference on history of medicine. Jenner’s papers are reprinted in Eliot (1910
David A. Freedman

Bazin (2000) discusses the history of smallpox, Jenner’s work, and later developments, including the eradication of smallpox; the last recorded cases were in 1977–78. There is a wealth of additional information on the disease and its history in Fenner et al. (1988).

Inoculation was recorded in England by 1721 (Bazin 2000: 13; Fenner et al. 1988: 214–6). However, the practice was described in the journals some years before that (Timonius and Woodward 1714). It was a common opinion in Jenner’s time that cowpox created immunity to smallpox (Jenner 1801; Baron 1838: 122). Over the period 1798–1978, techniques for producing and administering the vaccine were elaborated. As life spans became longer, it became clear that—contrary to Jenner’s teachings—the efficacy of vaccination gradually wore off. Revaccination was introduced. By 1939, the virus in the vaccines was a little different from naturally-occurring cowpox virus. The virus in the vaccines is called “vaccinia” (Bazin 2000: chap. 11; Fenner et al. 1988: chaps. 6–7, esp. 278).

Bulloch (1938) reviews the history of bacteriology. Bacteria were observed by Hooke and Leeuwenhoek before 1700. Otto Friderich Müller in Denmark developed a workable classification before 1800, improved about 50 years later by Ferdinand Cohn in Germany.

Some of Koch’s work on anthrax was anticipated by Pierre François Rayer and Casimir-Joseph Davaine in France. Likewise, Pasteur’s experiments disproving spontaneous generation built on previous work by others, including Lazzaro Spallanzani; contemporaneous research by John Tyndall should also be mentioned.

Freedman (2005: 6–9) reports on Snow and cholera. For detailed information on Snow’s work, see Vinten-Johansen et al. (2003). Evans (1987) gives a historical analysis of the cholera years in Europe. Koch’s discovery of the vibrio was anticipated by Filippo Pacini in 1854, but the implications of Pacini’s work were not recognized by his contemporaries.

Henry Whitehead was a clergyman in the Soho area. He did not believe that the Broad Street pump—famous for the purity of its water—was responsible for the epidemic. He saw a gap in Snow’s argument: the fatalities cluster around the pump, but what about the population in general? Whitehead made his own house-to-house canvass to determine attack rates among those who drank water from the pump and those who did not. Then he drew up a $2 \times 2$ table to summarize the results. The data convinced him that Snow was correct (Cholera Inquiry Committee 1855: 121–33). Snow made this kind of analysis only for his natural experiment.

William Farr, statistical superintendent of the General Register Office, was a leading medical statistician in Victorian England and a “sanitarian,” committed to eliminating air pollution and its sources. He claimed that the force of mortality from cholera in an area was inversely related to its eleva-
On Types of Scientific Inquiry

More specifically, if \( y \) is the death rate rate from cholera in an area and \( x \) is its elevation, Farr proposed the equation

\[
y = \frac{a}{b + x}
\]

The constants \( a \) and \( b \) were estimated from the data. For 1848–49, the fit was excellent.

Farr held the relationship to be causal, explained by atmospheric changes, including attenuation of noxious exhalations from the Thames, changes in vegetation, and changes in the soil. After the London epidemic of 1866, however, he came to accept substantial parts of Snow’s theory—without abandoning his own views about miasmas and elevation (Humphreys 1885: 341–84; Eyler 1979: 114–22; Vinten-Johansen et al. 2003: 394).

For better or worse, Farr’s belief in mathematical symbolism had considerable influence on the development of research methods in medicine and social science. Furthermore, the tension between the pursuit of social reform and the pursuit of truth, so evident in the work of the sanitarians, is still with us.

There are two informative web sites on Snow, Whitehead, and other major figures of the era (these sites were active as of January 8, 2010):

http://www.ph.ucla.edu/epi/snow.html
http://johnsnow.matrix.msu.edu/index.php

Loudon (2000) is highly recommended on puerperal fever; but also see Nuland (1979) for a more sympathetic account of Semmelweis’ life. Hare (1970: chap. 7) discusses efforts to control puerperal fever in a London maternity hospital in the 1930s. The strain of *Staphylococcus pyogenes* causing the disease turned out to be a common inhabitant of the human nose and throat (Loudon 2000: 201–4).

A definitive source on beriberi, only paraphrased here, is Carpenter (2000). He gives a vivid picture of a major scientific advance, including discussion of work done before Eijkman arrived in Java.

The discussion of pellagra is based on Freedman, Pisani, and Purves (2007: 15–16). Goldberger’s papers are collected in Terris (1964). Goldberger (1914) explains the reasoning that led him to the deficiency-disease hypothesis; Goldberger et al. (1926) identifies the P-P factor as part of the vitamin B complex. Carpenter (1981) reprints papers by many pellagra researchers, with invaluable commentary. He explains why in Mexico a corn-based diet does not lead to pellagra, discusses the role of tryptophan (an amino acid that can be converted to niacin in the body), and points
out the gaps in our knowledge of the disease and the reasons for its disappearance.

An excellent source on Fleming is Hare (1970), with Goldsmith (1946) adding useful background. Today, “penicillin” refers to the active ingredient in Fleming’s mold broth filtrate. What is the cell-killing mechanism? In brief, cell walls of most bacteria include a scaffolding constructed from sugars and amino acids. Components of the scaffolding have to be manufactured and assembled when the cells are dividing to form daughter cells. In many species of bacteria, penicillin interferes with the assembly process, eventually causing the cell wall to collapse (Walsh 2003).

Some species of bacteria manufacture an enzyme (“penicillinase”) that disables penicillin—before the penicillin can disable the cell. There are other bacterial defense systems too, which explain the limits to the efficacy of penicillin. Penicillin inhibits cell wall synthesis by a process that is reasonably well understood, but how does inhibition cause lysis? That is still something of a mystery, although much has been learned (Walsh 2003: 41; Bayles 2000; Giesbrecht et al. 1998).

Penicillin only causes lysis when bacteria are dividing. For this reason among others, a rather unusual combination of circumstances was needed to produce the effect that Fleming noticed on his Petri dish (Hare 1970: chap. 3). Was Fleming merely lucky? Pasteur’s epigram is worth remembering: “Dans les champs de l’observation, le hasard ne favorise que les esprits préparés.”

Almroth Wright, Fleming’s mentor, was one of the founders of modern immunology (Dunnill 2001). Among other accomplishments, he developed a vaccine that prevented typhoid fever. Wright was a close friend of George Bernard Shaw’s, and was the basis for one of the characters in The Doctor’s Dilemma.

---

1. This may be liberally translated as, “In the practice of observation, chance favors only the prepared mind.”
Data-set observations (DSOs) and causal-process observations (CPOs) provide two alternative foundations for causal inference. DSOs are the familiar data set of quantitative scholars, and research based on such data involves standard regression techniques and numerous variants on regression. CPOs, by contrast, are diagnostic “nuggets” of data that make a strong contribution to causal inference. The search for CPOs is a form of detective work that we call process tracing (Bennett, chap. 10, this volume), which seeks to establish the “physical and social processes through which purported causes affect outcomes” (Bennett and George 1997b: 3).

This chapter critically evaluates a study based on DSOs, and compares its conclusions with those of an analysis of CPOs.\(^1\) The substantive objective is to resolve one of the many controversies over the 2000 presidential election in Florida: the disputed outcome in the Florida Panhandle, which unlike the rest of the state is on Central Time. This led to a difference in timing that was at the center of the dispute.

---

1. For definitions of these two types of observations, see chapter 9 and the glossary in the present volume.
THE OPTION OF DSOs AND
REGRESSION ANALYSIS

John R. Lott argues that in the 2000 U.S. presidential election, at least 10,000 votes were lost for George W. Bush in the ten panhandle counties of Florida. The votes were lost because the networks declared Al Gore the winner in Florida after the polls had closed in eastern Florida but before the polls had closed in the panhandle counties, which are on Central Standard Time. Lott’s conclusion was widely discussed in the aftermath of the 2000 election and led to a series of congressional hearings.

To get his result, Lott employed a “difference-in-differences” form of regression analysis, based on data-set observations. He obtained turnout data on all sixty-seven Florida counties for four presidential elections (1988, 1992, 1996, and 2000), and he estimated a time-series cross-sectional regression with fixed county and time effects and with a “dummy variable” for the ten panhandle counties. In effect, Lott looked at the difference between one set of counties that got a “treatment” in the year 2000 (the ten panhandle counties whose polls were still open when the election was “called”) and those that did not (the remaining fifty-seven Florida counties in the eastern time zone), while controlling for differences reflected in the data from previous elections. Lott (2000) concluded that:

By prematurely declaring Gore the winner shortly before polls had closed in Florida’s conservative western Panhandle, the media ended up suppressing the Republican vote. . . . An examination of past Republican presidential votes by county in Florida from 1988 to 2000 shows that while total votes declined, the Republican voting rate in the western panhandle was significantly suppressed relative to the non-Republican vote. The 4 percent greater reduction in Republican votes averages about 1,000 votes per county, [yielding] 10,000 Republican votes for all 10 counties in the western Panhandle. This holds true even after accounting for the average differences in voting rates across counties as well as the changes in voting rates from one election to another.

2. This discussion is based on three sources. The first is Lott’s article in the November 14, 2000, Philadelphia Inquirer (Lott 2000) in which he provides a general description of his methodology and claims that 10,000 votes were lost. Second, Lott’s econometric analysis is described in Mason, Frankovic, and Jamieson (2001: 77–78). Third, Congressman Billie Tauzin subsequently held hearings on the elections and collected different analyses and interpretations of the vote. Congressman Tauzin’s office provided me with an annotated computer printout of Lott’s analysis, which reflects a methodology identical to that described both in Lott’s article and in Mason et al.

3. “Difference-in-differences” estimators are widely used in economics, and they are now a staple of introductory econometrics textbooks such as Stock and Watson (2003: 385–88) and Wooldridge (2009: chap. 13.2).
TURNING TO CAUSAL-PROCESS OBSERVATIONS

A researcher accustomed to the exclusive use of data-set observations might stop at this point, convinced that an adequate inference had been made. However, researchers oriented toward the use of causal-process observations would ask whether the result makes any sense. Is Lott’s estimate reasonable, given the number of voters who had not yet voted when the media called the election for Gore? How many of these voters heard the call? Of these, how many decided not to vote? And of those who decided not to vote, how many would have voted for Bush? Researchers can obtain answers to these questions by consulting diverse data sources and constructing a more intricate characterization of events on election day.

An inquiry to the networks established that the media calls were made ten minutes before the panhandle polls closed at 7:00 p.m.—twelve hours after the opening time of 7:00 a.m. If we assume that voters go to the polls at an even rate throughout the day, then only 1/72nd (ten minutes over twelve hours) of the voters had not yet voted when the media call was made. Alternatively, an analysis of Census data from 1996 on time of voting suggests that no more than about one-twelfth of the voters in Florida come to the polls in the last hour. If we assume that voters go to the polls at an even rate in this last hour, then (once again) only 1/72nd (one-sixth of one hour times one-twelfth) of the voters had not yet voted when the media call was made. Of the 379,000 voters in the panhandle, about 20 percent were absentee voters—leaving about 303,000 voters who voted on election day. One seventy-second of this figure is, in round numbers, 4,200 voters. The major assumption in this calculation is that voters come to the polls uniformly during the day or during the last hour. Interviews with Florida election officials and a review of media reports suggest that, typically, no rush to the polls occurs at the end of the day in the panhandle.

Only 4,200 people could have been swayed by the media call of the election, if they heard it. How many heard it? Research on media exposure suggests that an audience of 20 percent of adults for all media outlets would be very large. To be very conservative, I will assume that 20 percent of the 4,200 voters who intended to vote in the last ten minutes, or 840 people, heard the early call—though this is undoubtedly an overestimate because not all media were reporting the elections. Moreover, many of these prospective voters were Democrats or Independents who would not have voted for George W. Bush. In the panhandle, the Bush vote was about two-thirds of the total. If we assume the same proportion among those who were still to vote, it yields a total of 560 Bush voters who might have been affected.

Of these 560 Bush voters who might have heard the media call, how many decided not to vote? A review of past work on the impact of early
calls (Jackson 1983) and a general knowledge of voting behavior suggests a figure of 10 percent for the fraction of voters who decided not to vote once they knew the call was made for the presidential election. After all, voters select other officials as well, and they vote for reasons other than the likelihood that their vote will be decisive. Ten percent of 560 yields fifty-six Bush voters who might have been deterred from voting.

This estimate of Bush’s vote loss still probably exceeds the actual net effect. It seems just as likely that a Gore voter, rather than a Bush voter, might have decided not to vote. After all, for both candidates, the vote is no longer relevant to the presidential election once the call has been made. If 10 percent of the 280 Gore voters did not vote, then the net effect would be 28 Bush votes—56 Bush voters minus 28 Gore voters. This suggests a range of 28 to 56 Bush votes lost depending upon whether Gore voters were affected by the call. Even if we forget the offset for Gore voters and quadruple the estimate of 56 Bush voters who might have decided not to vote, the resulting upper-bound estimate of 224 voters is far short of the 10,000 that Lott claims.

My detective work leads to the inference that the approximate upper bound for Bush’s vote loss was 224 and that the actual vote loss was probably closer to somewhere between 28 and 56 votes. Lott’s figure of 10,000 makes no sense at all. This simple case-study analysis based upon information that goes beyond the turnout data used in the difference-in-differences model suggests a figure that is two orders of magnitude smaller than Lott’s result.

Although this case study of late voting uses quantitative data, it employs inferential tools typically associated with qualitative research. It draws upon multiple sources of information, utilizing inferences based on common sense, to establish an argument. It tries to approach the problem in several different ways, cross-checking information at every turn, and asking if the posited causal effect is probable, or even possible, given what we know from many different sources. In short, it investigates causal processes in close detail, and it tries to get beyond the results of an elaborate quantitative analysis of data-set observations.

WHERE DID LOTT GO WRONG?

The difference-in-differences method is widely used in economics and other social science disciplines as a way to adjust observational data for confounding factors that can lead to incorrect inferences. In this case, the method assumes that turnout in 2000 can be predicted by turnout in past years after adjusting for idiosyncratic factors of two types: those factors that affect each county in the same way over the entire time period but vary
from county to county (county fixed effects), and those factors that affect all counties in a given year but vary over years (time fixed effects).

This method does badly when idiosyncratic factors vary both by county and over time. For example, in 2000, organized labor put significant effort into increasing turnout in Florida, and it seems likely that it put its effort into mobilizing Democratic voters. As a result, turnout would be increased, compared to prior years, in counties with more Democrats (namely those outside the panhandle). The difference-in-difference method would not control for this. In fact, it would presume that the higher turnout outside the panhandle in 2000 should be translated into higher turnout inside the panhandle as well. To the extent that this higher turnout was not realized, Lott’s equation would pick it up as a negative coefficient on his dummy variable for the panhandle counties that he interpreted as the effect of the early media call. Instead, his coefficient might simply reflect labor’s success in mobilizing voters outside the panhandle.

In addition, quantitative methods are most believable when researchers are conservative about their inferences. Instead of using the standard .05 level of significance, Lott chose to use a .10 level, and he chose to employ a one-sided test that made his t-statistic of 1.285 just significant at this 10 percent level. This lenient approach to hypothesis testing allowed him to claim that his regression detected a significant effect. However, if Lott had decided to provide a 10 percent one-sided confidence interval for his estimate instead of a point estimate of 10,000, his confidence interval would have gone from zero to 20,000, thus providing little confidence in his assertions.

Even if these problems in Lott’s analysis were cleaned up by getting data on labor union activity and other factors, the analysis of such data would not necessarily supercede the inference based on causal-process observations. Even after putting aside the practical problems of collecting suitable data, it would be hard to collect data that could rule out all of the possible confounding effects. Consequently, rather than seeking additional data-set observations, in my judgment it would be more productive to do further in-depth analysis of causal-process observations drawn from these ten Florida panhandle counties, finding out what happened there, for example, by interviewing election officials and studying media reports.

CONCLUSION

Causal-process observations show that it was highly implausible for the media effect suggested by Lott’s analysis to have occurred. From a technical perspective, CPOs might be seen as a less sophisticated tool of analysis, yet
they effectively demonstrate that Lott’s quantitative conclusions based on regression analysis cannot be valid.

This chapter thus seeks to demonstrate the value of causal-process observations in what could be seen as a “least-likely case,” that is, a data-rich domain of mass political behavior. Even in this domain, this strategy of causal assessment provides valuable inferential leverage that supplements, and in this instance contradicts, the conclusions based on the analysis of data-set observations. Indeed, the lesson for quantitative researchers is the necessity of paying attention to the causal processes underlying behavior. Otherwise, regression analysis is likely to jump off the rails.
Addendum: Teaching Process Tracing

David Collier

Quantitative researchers receive extensive training in the spectrum of statistical tools employed in their research. By contrast, notwithstanding extensive efforts to institutionalize training in qualitative methods, techniques such as process tracing are not adequately taught. This deficit has motivated us to incorporate in this volume the three prior chapters on process tracing. Looking beyond these chapters, those concerned with graduate training in methodology should devote more systematic attention to process tracing. Given this need, we have included with the four online chapters on the Rowman & Littlefield website a set of exercises for teaching these analytic tools. To reiterate, this online material can be accessed using the instructions on the copyright page of this book.

1. Notable among these efforts are the Institute for Qualitative and Multi-Method Research, held annually at Syracuse University, and the Qualitative and Multi-Method Organized Section of the American Political Science Association. In addition to sponsoring panels at the annual meetings, the section offers short courses on qualitative methods. These trends are discussed in Collier and Elman (2008).
F. QUANTITATIVE TOOLS
FOR CAUSAL INFERENCE
Understanding Process Tracing

David Collier, University of California, Berkeley

ABSTRACT

Process tracing is a fundamental tool of qualitative analysis. This method is often invoked by scholars who carry out within-case analysis based on qualitative data, yet frequently it is neither adequately understood nor rigorously applied. This deficit motivates this article, which offers a new framework for carrying out process tracing. The reformulation integrates discussions of process tracing and causal-process observations, gives greater attention to description as a key contribution, and emphasizes the causal sequence in which process-tracing observations can be situated. In the current period of major innovation in quantitative tools for causal inference, this reformulation is part of a wider, parallel effort to achieve greater systematization of qualitative methods. A key point here is that these methods can add inferential leverage that is often lacking in quantitative analysis. This article is accompanied by online teaching exercises, focused on four examples from American politics, two from comparative politics, three from international relations, and one from public health/epidemiology.

Process tracing is a fundamental tool of qualitative analysis. In the framework presented here, it is defined as the systematic examination of diagnostic evidence selected and analyzed in light of research questions and hypotheses posed by the investigator. Process tracing can contribute decisively both to describing political and social phenomena and to evaluating causal claims. George and Bennett have played the leading role in developing this method as an essential form of within-case analysis, and Fenno’s “soaking and poking” is a kindred research procedure.

Although the idea of process tracing is often invoked by scholars as they examine qualitative data, too often this tool is neither well understood nor rigorously applied. Relatedly, the field of qualitative methods in political science—in sharp contrast to quantitative methods—is inadequately equipped with procedures for teaching basic research tools, including process tracing.

This two-fold deficit motivates this article, which offers a new framework for understanding, applying, and teaching process tracing. The approach is distinctive in three ways.

Process Tracing vis-à-vis CPOs. The evidence on which process tracing focuses corresponds to what Collier, Brady, and Seawright (2010a) call causal-process observations, or CPOs. The idea of CPOs highlights the contrast between (a) the empirical foundation of qualitative research, and (b) the data matrices analyzed by quantitative researchers, which may be called data-set observations (DSOs). Some of the literature on which this article draws (e.g., Brady 2010; Freedman 2010a; Mahoney 2010) formulates arguments in terms of CPOs, rather than in terms of process tracing per se. The present article treats these methodological tools as two facets of the same research procedure. Throughout, the article consistently refers to “process tracing” to avoid applying two labels to what is basically the same method.

Description. Careful description is a foundation of process tracing, a perspective emphasized by Mahoney (2010, 125–31). Process tracing inherently analyzes trajectories of change and causation, but the analysis fails if the phenomena observed at each step in this trajectory are not adequately described. Hence, what in a sense is “static” description is a crucial building block in analyzing the processes being studied.

Sequence. Process tracing gives close attention to sequences of independent, dependent, and intervening variables. Again, we follow Mahoney, who has productively advanced this approach.

TEACHING EXERCISES

This new formulation of process tracing is accompanied by online teaching exercises that encompass diverse substantive areas.


b. Comparative Politics: Lerner (1958) on social change in a Turkish village; and Rogowski (2010) on the interaction of theory and case studies.


David Collier is Chancellor’s Professor of Political Science in the Department of Political Science at the University of California, Berkeley. He can be reached at dcollier@berkeley.edu.
**The Teacher: Understanding Process Tracing**

**d. Public Health:** Freedman (2010a) on major breakthroughs in the history of epidemiology.

**e. Detective Fiction:** The Sherlock Holmes story “Silver Blaze” (posted online with the exercises) serves as the basis for an exercise, and also as a running example in parts of the following presentation. This story is not social science, yet it provides vivid illustrations of process tracing and is an engaging text for teaching.

**PROCESS TRACING, PRIOR KNOWLEDGE, AND DIAGNOSTIC EVIDENCE**

Process tracing, to reiterate, is an analytic tool for drawing descriptive and causal inferences from diagnostic pieces of evidence—often understood as part of a temporal sequence of events or phenomena. Given the close engagement with cases and the centrality of fine-grained case knowledge, process tracing can make decisive contributions to diverse research objectives, including:

1. Identifying novel political and social phenomena and systematically describing them;
2. Evaluating prior explanatory hypotheses, discovering new hypotheses, and assessing these new causal claims;
3. Gaining insight into causal mechanisms; and
4. Providing an alternative means—compared with conventional regression analysis and inference based on statistical models—of addressing challenging problems such as reciprocal causation, spuriousness, and selection bias. Thus, qualitative tools can add leverage in quantitative analysis. They can also strengthen causal inference in small-N designs based on the matching and contrasting of cases—designs which have great value, but whose contribution to causal inference urgently needs to be supplemented by within-case analysis.

Process tracing requires finding diagnostic evidence that provides the basis for descriptive and causal inference. How does the researcher establish that a given piece of evidence is diagnostic?

Identifying evidence that can be interpreted as diagnostic depends centrally on prior knowledge. For the purpose of the online exercises, we distinguish four interrelated types of knowledge—extending distinctions offered by Waltz (1979), whose ideas are important in the international relations examples included here.

**Conceptual Frameworks.** A first type of prior knowledge involves sets of interrelated concepts, often accompanied by general ideas of how the concepts can be operationalized. These frameworks thereby identify and link the topics seen as meriting analytic attention. The framework often points to the counterfactuals that conceptually establish what it means for a given phenomenon to be absent, that is, the “contrast space” (Garfinkel 1981) that organizes the analysis.

**Recurring Empirical Regularities.** These are established patterns in the relationships among two or more phenomena. Waltz (1979, 1) states that this is “not simply ... a relationship that has been found, but ... one that has been found repeatedly.” The corresponding “if a, then b” (1979, 1) connection may be viewed as causal, or it may be understood descriptively.

**Theory-I.** This builds on these recurring regularities by more tightly connecting them as a set of insights into “a particular behavior or phenomenon” (Waltz 1979, 2). Thus, many social scientists seek to build theory “by collecting carefully verified, interconnected hypotheses.”

**Theory-II.** A final type of prior knowledge entails not only interconnected empirical regularities (Theory-I), but also a set of statements that explain them, that is, offering explanations of why these regularities occur (Waltz 1979, 5). Theory-II may also be called an explanatory model.

As is clear in the exercises, some studies are explicit and precise about the prior knowledge that frames the research, whereas for other studies it is necessary to consult a wider literature to understand the theoretical background. Unfortunately, as investigators write up their research, they may overstate the coherence of the findings vis-à-vis prior knowledge—sometimes making it hard to identify the theoretical starting point. Reconstructing this starting point can require detective work—which is sometimes needed in evaluating diagnostic evidence in some of the exercises.

Against this backdrop, we consider the contribution of process tracing to descriptive and causal inference.

**DESCRIPTIVE INFERENCE**

Careful description is fundamental in all research, and causal inference—whether assessed with qualitative or quantitative tools—depends on it. Close engagement with case knowledge in process tracing can provide a good foundation for addressing this task.

A key point must be underscored again. As a tool of causal inference, process tracing focuses on the unfolding of events or situations over time. Yet grasping this unfolding is impossible if one cannot adequately describe an event or situation at one point in time. Hence, the descriptive component of process tracing begins with observing change or sequence, but rather with taking good snapshots at a series of specific moments. To characterize a process, we must be able to characterize key steps in the process, which in turn permits good analysis of change and sequence.

Mahoney (2010, 127–28) illustrates descriptive inference in process tracing with Tannenwald’s (1999) study of the “Nuclear Taboo.” Tannenwald argues that the horrified reaction to the use of nuclear weapons at the end of the World War II created a nuclear taboo that strongly influenced later US nuclear policy, specifically decisions about the non-use of nuclear weapons during subsequent military crises. Whereas this taboo grew out of the reaction at the level of public opinion, it evolved into a normative mandate embraced by policymakers (1999, 462). A crucial task in Tannenwald’s study is to establish empirically (a) that this horrified reaction did in fact occur; (b) how widespread it was; and (c) that the elements of this reaction did indeed add up to a nuclear taboo. Process tracing focuses on finding and interpreting diagnostic evidence that addresses these descriptive tasks. This nuclear taboo, in turn, is the key independent variable in the study that is evaluated vis-à-vis rival explanations of the non-use of nuclear weapons.

Lerner’s (1958) analysis of rapid “modernization” in a Turkish village likewise illustrates the intensive description that should be a foundation of process tracing. This transformation results from the election of a new national governing party and the subsequent introduction of infrastructure that includes electricity and a modern road to Ankara. The transformation of the village is the dependent variable, and the author’s goal is to describe change in this variable over time. The analysis focuses on dozens of specific observations of social attributes and interactions; demographic characteristics; and material objects, physical infrastructure, and commercial establishments.
These two examples of description differ in important ways. Tannenwald relies on diverse primary and secondary sources—including official documents, memoirs, and biographies—that shed light on the politics of nuclear policy making. By contrast, Lerner’s study depends on intensive interviewing, carried out by his field assistants. Further, as noted, the phenomenon described through process tracing is Tannenwald’s main independent variable; thus, the nuclear taboo is her hypothesized explanation for the US post-World War II non-use of nuclear weapons. By contrast, for Lerner the posited modernization of the village is the dependent variable triggered by an electoral shift and the initiative of the victorious party to build new infrastructure. Finally, while both Lerner and Tannenwald offer a rich and detailed description of a key variable, Tannenwald also gives substantial attention to rival explanations.

The two examples also illustrate another point: The qualitative researcher should recognize that the fine-grained description in process tracing sometimes relies on quantitative data. This is certainly reasonable, given that—in the spirit of pursuing multi-method research—the boundary between qualitative and quantitative should not be rigid. For Lerner, some of the information is demographic, involving numerical data. As seen in the exercises, Brady’s (2010) process tracing study employs quantitative data on elections and voting. In parallel, Tannenwald could have assessed the perversiveness of horrified reactions by counting their overall frequency, different types of horrified reactions, and change in these counts over time. Process tracing does indeed focus on single “nuggets” of information, yet sometimes this information involves counts and not just single actions or occurrences.

A different form of description, based on counterfactuals, is illustrated by the Sherlock Holmes story “Silver Blaze.” Here the central puzzle is to explain the murder of John Straker, trainer of the racehorse Silver Blaze. The focus is on a singular event that cannot be disaggregated—a focus also common in process-tracing research in international relations. With singular events, description may be based on comparison of the observed value of a given variable with one or more hypothetical—i.e., counterfactual—values that are seen as plausible alternatives, but that do not occur in the case being studied. The comparison depends on the contrast space noted above, which builds on the researcher’s background knowledge. Counterfactuals are important in diverse areas of research (e.g., King, Keohane, and Verba 1994, 77–78, 88–89), and they play a particularly visible role here.

**CAUSAL INFERENCE**

Basic ideas about applying process tracing to causal inference may be summarized in terms of four empirical tests. Slightly adapting the formulation of Bennett (2010), who builds on the work of Van Evera (1997), the tests are classified according to whether passing the test is necessary and/or sufficient for accepting the inference. Based on these criteria, table 1 presents the four tests: straw-in-the-wind, hoop, smoking-gun, and doubly decisive. The table also notes the implications for rival hypotheses of passing each test. If a given hypothesis passes a straw-in-the-wind test, it only slightly weakens rival hypotheses; with hoop tests it somewhat weakens them; with smoking-gun tests it substantially weakens them; and with doubly decisive tests passing eliminates them—of course, with the usual caveat that the definitive elimination of a hypothesis is often hard to achieve in social science.

Before we introduce causal inference, it is useful to reiterate two ideas discussed above: process tracing can focus either on recurring events or on a singular event; and although it is reasonable to think of process tracing as a qualitative method, it sometimes relies on quantitative information. Three other points should also be emphasized:

**Specification of Hypotheses.** Careful, analytically informed specification of hypotheses is essential both in selecting and interpreting pieces of evidence, and in weighing them against one another. Background knowledge is fundamental here.

**Distinctions among Tests.** The distinctions in table 1 are a useful heuristic, but should not be taken rigidly. The decision to treat a given piece of evidence as the basis for one of the four tests can depend on the researcher’s prior knowledge, the assumptions that underlie the study, and the specific formulation of the hypothesis. Although in general the appropriate test is clear, sometimes a piece of evidence treated as a straw-in-the-wind might instead be viewed as the basis for a hoop test or a smoking-gun test (see tables 4 and 5 below). Alternatively, it might simply be viewed as an “intermediate” test, with corresponding implications for rival hypotheses.

---

**Table 1**  
**Process Tracing Tests for Causal Inference**

<table>
<thead>
<tr>
<th>Necessary for Affirming Causal Inference</th>
<th>Sufficient for Affirming Causal Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>a. Passing: Affirms relevance of hypothesis, but does not confirm it</td>
<td>a. Passing: Confirms hypothesis.</td>
</tr>
<tr>
<td>b. Failing: Hypothesis is not eliminated, but is slightly weakened</td>
<td>b. Failing: Hypothesis is not eliminated, but is somewhat weakened</td>
</tr>
<tr>
<td>c. Implications for rival hypotheses: Passing: slightly weakens them</td>
<td>c. Implications for rival hypotheses: Passing: substantially weakens them</td>
</tr>
<tr>
<td>2. Hoop</td>
<td>4. Doubly Decisive</td>
</tr>
<tr>
<td>a. Passing: Confirms hypothesis and eliminates others</td>
<td>a. Passing: Confirms hypothesis and eliminates others</td>
</tr>
<tr>
<td>c. Implications for rival hypotheses: Passing: eliminates them</td>
<td>c. Implications for rival hypotheses: Passing: eliminates them</td>
</tr>
<tr>
<td>Failing: somewhat strengthens them.</td>
<td>Failing: substantially strengthens them.</td>
</tr>
</tbody>
</table>

Table 2
Overview of “Silver Blaze”

<table>
<thead>
<tr>
<th>Causal Puzzle</th>
<th>Intervening Variables</th>
<th>Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>To explain the murder of John Straker and, secondarily, the disappearance and whereabouts of the racehorse Silver Blaze.</td>
<td>H3. Straker abducted horse</td>
<td>H6. Simpson killed Straker</td>
</tr>
<tr>
<td>H1. Romantic entanglement started chain of events</td>
<td>H4. Straker planned to harm horse</td>
<td>H7. Straker killed himself</td>
</tr>
<tr>
<td>H2. Chain of events started in Straker household</td>
<td>H5. Straker practiced the injury</td>
<td>H8. Horse killed Straker</td>
</tr>
</tbody>
</table>

**Main Characters**

Silver Blaze, the racehorse that is the favorite for the Essex Cup, has disappeared.

John Straker, the horse’s trainer, has been killed by a terrible blow that shattered his head.

Fitzroy Simpson, a prime suspect, has been lurking around the stable seeking inside information about the race.

Ned Hunter, a stable boy, has been drugged with opium concealed in curried mutton. He therefore fails to guard Silver Blaze on the night of the horse’s disappearance.

Colonel Ross is the owner of King’s Pyland Stables and of Silver Blaze.

**Hypotheses**

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Interfering Variables</th>
<th>Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2. Chain of events started in Straker household</td>
<td>H4. Straker planned to harm horse</td>
<td>H7. Straker killed himself</td>
</tr>
<tr>
<td></td>
<td>H5. Straker practiced the injury</td>
<td>H8. Horse killed Straker</td>
</tr>
</tbody>
</table>

**Assumptions and Interpretations.** The decision about which test is appropriate to a particular piece of evidence thus involves different assumptions and interpretations. For example, if researchers make the weaker assumption that a given event (or other piece of evidence) may be a coincidence, they should and will be more cautious. Alternatively, if they make the stronger assumption—based on prior knowledge—that it is probably not a coincidence, they may arrive at a different conclusion about accepting or rejecting the hypothesis.

Against this backdrop, we discuss the four process-tracing tests in table 1, using the Sherlock Holmes story “Silver Blaze” as a running example. At this point it may be useful for readers to examine the story itself.

Mapping the Holmes story onto the framework presented earlier, we might say that the suspects are in effect the hypotheses, and the clues are causal process observations (CPOs). Table 2 provides an overview of the story and presents the hypotheses used in illustrating the four tests, organized according to whether they concern an independent, intervening, or dependent variable. As indicated in the table, the mystery contains two causal puzzles: explaining the murder of John Straker and the disappearance of the horse. The following examples concentrate on the murder, and references to the horse’s disappearance are considered when crucial to the murder itself.

**Straw-in-the-Wind Tests**

These tests, illustrated in table 3, can increase the plausibility of a given hypothesis or raise doubts about it, but are not decisive by themselves. Straw-in-the-wind tests thus provide *neither a necessary nor a sufficient criterion* for accepting or rejecting a hypothesis, and they only *slightly* weaken rival hypotheses. Of the four tests, these are the weakest and place the least demand on the researcher’s knowledge and assumptions. Yet they provide valuable benchmarks in an investigation by giving an initial assessment of a hypothesis. Furthermore, if a given hypothesis passes multiple straw-in-the-wind tests, it adds up to important affirmative evidence.

In “Silver Blaze,” one straw-in-the-wind is based on the clues about the bill for expensive women’s clothing found in Straker’s pocket and Straker’s wife’s ignorance of the costly dress that had been purchased. This lends weight to Holmes’s suspicion about Straker’s role (H1) and to the idea that Straker might have had a financial motive for throwing the race, but is not by itself a decisive piece of evidence. Another straw-in-the-wind is one of the most famous clues in all of detective fiction: that the dog presumably guarding the horse’s stable “did nothing in the night,” an observation that points to the possibility that someone known to the dog—i.e., Straker—abducted the horse (H3). Yet it certainly does not confirm this hypothesis.

**Hoop Tests**

Hoop tests (table 4) set a more demanding standard. The hypothesis must “jump through the hoop” to remain under consideration, but passing the test does not by itself affirm the hypothesis. Although not yielding a *sufficient* criterion for accepting the explanation, it establishes a *necessary* criterion. Hoop tests do not confirm a hypothesis, but they can eliminate it. Compared to the straw-in-the-wind tests, passing hoop tests has stronger implications for rival hypotheses: it *somewhat* weakens their plausibility, without precluding the possibility that alternative hypotheses may be relevant.

Table 3
Straw-in-the-Wind Tests

| H1. | Straker’s romantic entanglement set chain of events into motion. |
| Clues | A bill from an expensive women’s clothing store is found in Straker’s pocket, and his wife is ignorant of the clothing in question. |
| Inference | The bill was owed by Straker for an expensive gift to another woman, and Straker may have been in financial difficulty. This could give him a motive for throwing the race. |
| Summary | This promising lead, a straw-in-the-wind, lends weight to H1, but is not by itself a decisive piece of evidence. |

| H3. | Straker abducted the horse. |
| Clue | The dog did nothing (i.e., did not bark) in the night during which the horse disappeared. |
| Inference | The person who approached the stable, possibly Straker, was well-known to the dog. This raises questions about why Straker might have gone to the stable. It suggests that perhaps he came to abduct the horse, but does not strongly demonstrate this. |
| Summary | This straw-in-the-wind favors H3, but does not confirm it. |
Table 4

Hoop Tests

<table>
<thead>
<tr>
<th>Clue</th>
<th>Inference</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>H6. Simpson killed Straker.</td>
<td>This weapon is consistent with the hypothesis, but does not by itself demonstrate Simpson’s guilt.</td>
<td>Simpson had a potential murder weapon, so H6 passes this hoop test.</td>
</tr>
<tr>
<td>Clue. Simpson had a potential murder weapon.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inference.</td>
<td>With a stronger assumption based on his appearance, Simpson could not have inflicted the blow that shattered Straker’s head.</td>
<td>Hence, he fails this second hoop test. Table 4 also illustrates alternative interpretations of the same piece of evidence. One could assume, as just stated, that a timid individual such as Simpson would never commit a savage murder—thereby eliminating him as a suspect through the hoop test. Alternatively, it may be unlikely, but definitely not impossible, that a timid and non-namening person would commit such a murder. This would yield a straw-in-the-wind test which fails; with a weaker assumption it is a straw-in-the-wind test which casts doubt on H6.</td>
</tr>
<tr>
<td>Summary.</td>
<td>With a stronger assumption this is a hoop test which H6 fails; with a weaker assumption it is a straw-in-the-wind test which casts doubt on H6.</td>
<td></td>
</tr>
</tbody>
</table>

Table 5

Smoking-Gun Tests

<table>
<thead>
<tr>
<th>Clue</th>
<th>Inference</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2. The chain of events started in Straker’s household.</td>
<td>The clues yield a smoking-gun test that confirms H2.</td>
<td></td>
</tr>
<tr>
<td>Clues. The household maid brought the stable boy curried mutton, and he was found later to have been drugged.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inference. The curry was served to conceal the opium, which in turn was used to drug the stable boy. When it is clear that the curry could only have been introduced in the mutton by someone in Straker’s household, members of his household become inextricably linked to a key causal step.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary.</td>
<td>If the knife is exceptionally unusual, it is a smoking gun that confirms H4. With a weaker interpretation that the knife was somewhat unusual, it is a straw-in-the-wind that makes H4 more plausible, without confirming it.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 presents two hoop tests, both focused on the hypothesis that Simpson killed Straker (H6). Simpson carried a potential murder weapon, so that he passes the corresponding hoop test and is therefore not precluded as a suspect. However, his timid, non-menacing appearance might seem to preclude his being a murderer who “shattered” Straker’s head with a “savage” blow. Hence, he fails this second hoop test. Table 4 also illustrates alternative interpretations of the same piece of evidence. One could assume, as just stated, that a timid individual such as Simpson would never commit a savage murder—thereby eliminating him as a suspect through the hoop test. Alternatively, it may be unlikely, but definitely not impossible, that a timid and non-namening person would commit such a murder. This would yield a straw-in-the-wind test that casts doubt on the idea that he is the murderer—yet he remains a possible suspect.

Smoking-Gun Tests

The metaphor of a “smoking gun” conveys the idea that a suspect who is caught holding a smoking gun is presumed guilty. However, those with no smoking gun may not be innocent. In other words, this provides a sufficient but not necessary criterion for accepting the causal inference. It can strongly support the given hypothesis, but failure to pass does not reject it. If a given hypothesis passes, it is substantially weakens rival hypotheses.

In “Silver Blaze,” the first smoking-gun test (table 5) is straightforward. The fact that the maid brought the curried mutton to the stable shows that the initiative to drug the stable boy—a key step in the chain of events—had to begin in Straker’s household (H2). By contrast, the hypothesis that Straker planned to cause harm (H4) is ambiguous and illustrates the importance of prior knowledge and assumptions. Depending on such knowledge and assumptions, the knife found with Straker can be viewed as extraordinarily odd and suspicious, or only somewhat unusual. Accordingly, the knife is alternatively a smoking-gun or a straw-in-the-wind.

Doubly Decisive Tests

These tests provide strong inferential leverage that confirms one hypothesis and eliminates all others. They meet both the necessary and sufficient standard for establishing causation. As Bennett (2010, 211) notes, single tests that accomplish this are rare in social science, but this leverage may be achieved by combining multiple tests, which together support one explanation and eliminate all others.

Turning again to the Sherlock Holmes example (table 6), we see that Simpson, Straker, and the horse are suspects in Straker’s death. Simpson (H6) and Straker (H7) are removed from suspicion by hoop tests. Based on straw-in-the-wind tests, one of which might be interpreted as a smoking-gun test, Holmes infers that the horse kicked Straker, thereby inflicting the grievous blow that shattered his head. The conjunction of these diverse tests serves to eliminate other suspects and establish the horse’s guilt (H8), thereby meeting the standard of both necessity and sufficiency.

Combining tests in this way poses important challenges. Central here is Holmes’s method of elimination, a strategy invoked various times in Doyle’s stories. Put simply, when the investigator has eliminated all plausible alternatives, the remaining scenario must be the correct one. Variants of this method are widely recognized, as with eliminative induction in Bayesian analysis (Vineberg 1996) and J.S. Mill’s (1974, 397–98) method of residues.

The method of elimination is especially relevant here because although two suspects are definitely eliminated through hoop tests, the guilt of the horse is established primarily on the basis of weaker straw-in-the-wind tests. The procedure of elimination is valuable because it relies centrally on the definitive exonerations of the first two suspects, and only secondarily on explicit demonstration of the horse’s guilt.

Further, the method of elimination has special relevance to the case of Silver Blaze, given that both Mill (1974, 398) and Holmes emphasize the value of this method for discovering unusual or even bizarre explanations—such as the kick of the horse as a murder weapon (see table 6, H8, inference d). Mill states that among his methods, “this is the most fertile in unexpected results” (p. 398), and as Holmes puts it, “when you have eliminated the impossible, whatever remains, however improbable, must be the
In contemporary political science, a recurring concern is that the discipline needs tools for discovering unexpected and unusual explanations. The method of elimination merits attention as precisely this kind of tool.

**Causal-Sequence Framework: Auxiliary Outcome Test**

Another test is suggested by Mahoney’s (2010, 125–31) causal-sequence framework. He introduces the idea of auxiliary outcomes, which are not part of the main causal sequence yet provide valuable inferential leverage. Fruitful theories generate multiple observable implications (King, Keohane, and Verba 1994, passim), and a particular independent variable or mechanism hypothesized to influence the dependent variable may also affect an auxiliary outcome. The inference thereby derived may further support the causal importance of the independent variable or intervening mechanism. In “Silver Blaze,” the lame sheep are an example of an auxiliary outcome (table 7). Holmes infers that Straker wished to practice the delicate operation required to injure the horse (H5). This does not directly injure the horse; it is a secondary outcome that makes Holmes’s ideas about the central causal process more plausible. Auxiliary outcomes generally yield straws-in-the-wind, as occurs in this example.

Mahoney further illustrates the auxiliary outcome test with a social science example: Luebbert’s (1991) famous book, Liberalism, Fascism, or Social Democracy. Luebbert’s central argument is that a “red-green” coalition of socialist parties and the middle peasantry was a key factor in the formation of national-political economies in interwar Europe. Mahoney shows that while this claim is partly developed through small-N comparative research and partly through a focus on mechanisms, Luebbert also builds his case by arguing that if a red-green alliance really did foster social democracy, it should have left behind other markers, including the reluctance of socialists to challenge the distribution of wealth in the countryside (Mahoney 2010, 130). The discovery of such auxiliary outcomes suggests that the red-green alliance had a key impact on other domains of national politics. This finding reinforces the idea that the alliance was highly influential, yielding stronger grounds for inferring that it also shaped the national political-economic regime (Mahoney 2010, 130).

**CONCLUSION**

This article seeks to improve the practice of process tracing as a strategy of qualitative analysis, a strategy that can also contribute to quantitative research. The discussion is accompanied by the online exercises focused on ten empirical studies, from diverse subfields, aimed at encouraging careful thinking—and productive teaching—about process tracing.

Three concluding points merit emphasis. First, as Brady, Collier, and Seawright (2010, 22) note, “both qualitative and quantitative research are hard to do well.” Qualitative tools such as process tracing can address some challenges faced in quantitative analysis, but process tracing faces serious issues in its own right. Doubts may arise as to which causal-inference test is appropriate. The analysis may face standard problems of missing variables. Measurement error can be an issue, and probabilistic relationships are harder to address than in quantitative research. This article is intended as one step in developing and refining tools for process tracing—and it is urgent that it not be the last step. More work must be done.

Second, in a given study, how does one begin to carry out process tracing? It is certainly valuable to approach process tracing with the expectation of using the causal inference tests presented in table 1, yet these tests are not always easy to apply. It can therefore be productive to start with a good narrative or with a timeline that lists the sequence of events. One can then explore the causal ideas embedded in the narratives, consider the kinds of evidence...
that may confirm or disconfirm these ideas, and identify the tests appropriate for evaluating this evidence.

Finally, along with the value per se of refining process tracing, this discussion is important in wider debates on political methodology. Political science is in a period of major innovation in refining tools for quantitative analysis, and in particular, quantitative tools for causal inference. This trend has produced some worries among qualitative researchers about the adequacy of their own tools, and perhaps it has intensified the skepticism of some quantitative researchers about causal inference in qualitative studies. This skepticism led the eminent statistician David Freedman (2010b) to counter with the argument that the kind of qualitative analysis involved in process tracing is indeed a type of scientific inquiry in its own right. In that spirit, the goal here is to take steps toward placing this form of inquiry on a more rigorous foundation.

NOTES
Among the several colleagues who provided valuable comments on this article, Maria Gould and three anonymous reviewers for PS deserve special thanks.

1. The approach discussed here differs from other research traditions that can be linked to the idea of process tracing—for example, the work on mechanisms of Tilly (2001) and McAdam, Tarrow, and Tilly (2001).

2. Within-case analysis can become multi-case analysis if different facets of the initial “case” are analyzed. The key idea here is that the point of departure is a single case, when viewed from the perspective of a wider comparative analysis focused on a larger N.


5. Ideas about these designs based on a matching of cases are often drawn from J.S. Mill (1974) and Przeworski and Teune (1970). For a comment on the weakness of these designs for causal inference, see Collier, Brady, and Seawright (2010a, 16).

6. Addressing this question raises issues about the logic of inquiry and the form of social scientific knowledge that are well beyond the scope of this discussion. Only a few basic points are addressed here that are salient for the accompanying exercises.

7. Obviously, such prior knowledge is essential in all research, both qualitative and quantitative.

8. Waltz calls claims about these regularities “law-like statements” (p. 1). We prefer the alternative label used here.

9. The expressions “descriptive inference” and “causal inference” are employed here in the sense of King, Keohane, and Verba (1994, 7–8, chapters 3–5). Their usage can be seen as approximating an ordinary language meaning of “description” and “causation”; and by “inference” they mean that researchers have “the goal of making inferences that go beyond the particular observations collected,” that is, they are analyzed within the larger framework used by the investigator. This usage contrasts with ideas of “descriptive inference” and “statistical inference” that are standard in the work of statisticians (e.g., Berk 2004, chapter 11).

10. Achieving good description in this sense, and developing fruitful ideas about the unfolding of the process, may of course interact in an iterative manner.

11. Tannenwald’s study is also discussed in Collier, Brady, and Seawright (2010a, 189–90; 2010b, 509).

12. Lerner’s analysis—which is the focus of one of the exercises—is closely tied to modernization theory, which might concern some readers; and at certain points the translation seems descending. Further, the female interviewer is presented in a sexist way (although in survey research, selecting interviewers in light of characteristics such as these is widely recognized as important). However, these drawbacks are outweighed by the opportunity presented by the chapter to illustrate the practice of making careful observations, and also to see how they can be integrated into a complex picture of social change.

13. For a framing of qualitative vis-a-vis quantitative in terms of four dimensions, see Collier, Brady, and Seawright (2010a, 177–82).

14. Another puzzle is explaining the disappearance of the horse, but as Holmes himself emphasizes, that is a secondary issue (see p. 13 in the accompanying online version of the story).


16. There is a parallel here to the idea in statistical work that the test does not stand on its own, but rather is shaped by prior assumptions. In quantitative analysis, the construction of the statistical model depends heavily on such assumptions, and in general the statistical test does not directly evaluate these assumptions. Rather, it estimates the relationship based on the supposition that the model assumptions, as well as the underlying assumption of causal- ity, are true. See, for example, Freedman (2010b).

17. In one story, Sherlock Holmes takes a strong stand on coincidences (“Adventure of the Second Stain”; in Doyle 1960, this is on p. 655). Watson refers to the juxtaposition of two key events as “an amazing coincidence.” Holmes replies: “A coincidence! The odds are enormous against its being a coincidence. No figures can express them. No, my dear Watson, the two events are connected—must be connected. It is for us to find the connection.” Ironically, it turns out that these two events are only tangentially connected, so Watson’s statement was closer to the truth than Holmes’s, and the weaker assumption was more appropriate. In another story, Holmes is initially more cautious about inferences and coincidences, but then he backtracks and insists on the certitude of his inferences (“The Sign of Four,” chap. 1; in Doyle 1960, this is on p. 93).

18. “The Sign of Four,” chap. 6 (in Doyle 1960, this is on p. 111).

19. A further perspective on unusual or bizarre explanations (see again table 6, H8, inference d) derives from William James’s famous dictum that “every difference must make a difference.” “To put this in a less extreme form, it might be said that some differences make a difference. In this instance, the form of the murder was so distinctive that it called for a distinctive explanation—which turned out to be the kick of a horse. On William James, see Copi (1955, 33–32).

20. Andrew Bennett (personal communication) has underscored the parallel here with diagnostic tests in medicine.


REFERENCES


The Teacher: Understanding Process Tracing


Taking Temporality Seriously: Modeling History and the Use of Narratives as Evidence
Author(s): Tim Büthe
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/3117924
Accessed: 08-07-2015 14:13 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp
JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.
Taking Temporality Seriously: Modeling History and the Use of Narratives as Evidence

TIM BÜTHE  Columbia University

Social scientists interested in explaining historical processes can, indeed should, refuse the choice between modeling causal relationships and studying history. Identifying temporality as the defining characteristic of processes that can be meaningfully distinguished as “history,” I show that modeling such phenomena engenders particular difficulties but is both possible and fruitful. Narratives, as a way of presenting empirical information, have distinctive strengths that make them especially suited for historical scholarship, and structuring the narratives based on the model allows us to treat them as data on which to test the model. At the same time, this use of narratives raises methodological problems not identified in recent debates. I specify these problems, analyze their implications, and suggest ways of solving or minimizing them. There is no inherent incompatibility between—but much potential gain from—modeling history and using historical narratives as data.

Across the empirical subfields of political science, we find in recent years a renewed and growing interest in “historical macro-analysis” (Katznelson 1997, 82), which seeks to understand and causally explain processes with an important temporal dimension, such as the formation and evolution of formal and informal social institutions. This trend has given rise to “historical institutionalism” (for overviews see Hall and Taylor 1996; Immergut 1998; Robertson 1993; Thelen 1999), which manifests itself in much of the newer literature on the welfare state and state formation in comparative politics and in the literature on American political development. In international relations, manifestations of this trend toward historical scholarship range from the interest in domestic and international institutions to the postmortem debate over the nature of the Cold War. Does this interest in historical processes merely expand the subject matter of political science or does it raise particular methodological problems that require a distinct approach to theorizing and to the presentation of empirical information to test the plausibility or validity of our explanations?1

To advance and extend the methodological debate on this general question, I examine two sets of specific issues. First, what defines “history” as a distinct object of study? What are the implications of such a conception of history for developing explicit theoretical models? Is there an inherent incompatibility between modeling and the quest for explanation and narration of “history,” as some observers suggest (e.g., Elster 2000)? Based on an inclusive conception of modeling and an explicit conception of “history” as processes rendered distinctive by the importance of temporality, I argue that modeling such processes is particularly difficult but, nonetheless, possible and desirable. Far from being inherently futile, modeling history is extremely useful, not least because models, by emphasizing the general, help us clarify what is historically and contextually specific when we examine the historical record. Consequently, the “historic turn in the human sciences” (McDonald 1996a) need not lead us away from what scholars of very different persuasions have identified as the particular strength and source of progress of American political science: the explicit modeling of the political phenomena we seek to explain, so as to facilitate scrutiny of the deductive logic of the explanation (see, e.g., Milner 1998, and Waever 1998).

Second, because most historical work in political science is narrative in form, I examine the relationship between models and narratives and, more generally, the strengths and weaknesses of narratives as a type of “presentation of the results” of our analysis (Skocpol 1995, 44). Here, my focus differs from much recent work that concentrates on narratives as a source of empirical information for the analyst, either broadly, conceptualizing all empirical information as “text” to be interpreted (Ricoeur [1971] 1979), or more narrowly, concentrating on specific oral and written “histories” in the form of narratives as constitutive elements of ideas and norms (e.g., Anderson 1983) 1991; Finnemore 1996; Finnemore and Sikkink 1998) with interesting implications for policy (e.g., Van Evera 1994, 36f). These works have been joined by predominantly methodological contributions (e.g., Heise 1993; Lustick 1996), which often seek to raise social scientists’ awareness of historians’ carefully developed methods to discern

---

1 Throughout this paper I subsume epistemological issues under “methodological” ones.
MODELING

Any attempt to answer the question of whether modeling and the study of history are incompatible should start with explicit definitions of modeling and history. I take a model to be a schematic statement of a theoretical argument, a hypothesized parsimonious abstraction or simplification of “reality” that depicts a deductively sound, systematic, regular relationship between specified aspects of reality and helps to explain that relationship. Modeling to provide causal explanations, we customarily start by designating the explanandum as the “dependent variable” and its hypothesized causes as the “independent variables.”

The historical narratives discussed here are therefore in many respects similar to case studies, discussed in a number of books and articles on methodology in recent years (e.g., Achen and Snidal 1989; Bennett 1999; Collier 1995; Fearon 1991; Geddes 1990; Jervis 1990; King, Keohane, and Verba 1994; Przeworski 1995; Rogowski 1995; Tilly 1997). The resulting models allow us, with the help of specified assumptions, to derive specific hypotheses about the explanandum, given particular “values” for the independent variables—though the specificity may be restricted to predicting the direction of change in the explanandum. Whether employing the logic of instrumental rationality or not, whether assuming perfect strategic and computational capabilities or more limited ones, whether working quantitatively or qualitatively, we model to emphasize what we consider to be the most important, systematic, and in that sense generalizable elements of the phenomena we seek to explain.

Taking Temporality Seriously

September 2002

428

This content downloaded from 130.132.173.87 on Wed, 08 Jul 2015 14:13:49 UTC
All use subject to JSTOR Terms and Conditions
increase clarity and explicitness, reducing ambiguity and vagueness. That said, in practice, models often are not explicit about all elements of the analysis. Particularly when one mode of analysis dominates a discipline, its core assumptions are often dropped from discussion. These assumptions remain a central part of the analysis, but they are no longer mentioned and may be “smuggled in” and thus become “immune to rational criticism” (Barry 1985, 282). Yet, although modeling can be a rhetorical device to hide one’s values or other assumptions (McCloskey 1998), it need not be; and explicit modeling surely makes it easier to detect hidden assumptions than leaving the theoretical argument implicit.

Modeling as such, then, does not require any particular ontological, epistemological, or substantive assumptions, though any specific model will have to make them and should make them explicit to facilitate their empirical as well as their analytical scrutiny. By definition, assumptions must be “unrealistic” or “inaccurate” in the sense of being incomplete. But if assumptions are manifestly empirically wrong—as in Milton Friedman’s ([1953] 1979, 30) (in)famous example of the leaf-growth pattern on a tree—they cause two major problems. First is the familiar problem of the joint hypothesis test: It will be impossible to tell whether empirical disconfirmation of a hypothesis derived from a model with such assumptions is due to a flaw or omission in the model or due to the erroneous assumptions. Second, when predictions that are based on manifestly wrong assumptions are confirmed empirically, such findings are not very useful, because the models based on such assumptions provide little insight into how the outcomes came about (Coase [1981] 1994, 17). Policy that seeks to achieve or avoid these outcomes, if based on such models, is likely to be ineffective and might be outright dangerous.

The final point to be made before turning to a definition of “history” concerns the roles of deduction and induction in modeling. Deduction is of central importance for the ex ante identification of the elements—the types of (f)actors—that constitute the building blocks of a possible model. It is also the central heuristic tool for formulating logically sound relationships among these elements, consistent with the model’s core ontological assumptions and consistent with the assumptions driving the causal mechanisms, which themselves are usually derived from very general theories of the constraints, motivations, and cognitive processes employed in decision making and thus shaping human agency. But a model based solely on deduction from assumptions is a shell consisting of “empty theoretical boxes” (Bonnell 1980, 162; see also Smelser 1967, 22ff, 32); it needs empirical content in order to make predictions about what we should observe in a particular instance. To be sure, this content could be supplied by further assumptions that are only subsequently subjected to empirical scrutiny (Kiser and Hechter 1998, 790ff, 802ff), but if we truly had no empirical knowledge at all, what would be the basis for such conjectures? In practice, therefore, the empirical content will probably be supplied inductively. Moreover, induction often has a further role in the modeling process, especially when empirical disconfirmation of our initial model leads us to the inductive discovery of additional explanatory factors that can account for the anomalies. Integrating these factors systematically into the model is surely preferable to having them remain extraneous and theoretically irrelevant (cf. Smelser 1967, 33, 35), as long as we keep in mind that the inductive addition of “auxiliary hypotheses” without subsequent separate tests leads only to hypotheses, not conclusions (King, KEOHANE, and Verba 1994, 214, 101ff; Lakatos [1970] 1974, 117f). That said, even works that are entirely inductive would benefit in two ways from making their argument explicit and separating it as a model from the particular case on which it is based. First, such a separation forces the author to distinguish between the conceptually abstract elements of causal relationships and their particular manifestations. This distinction is a prerequisite for assessing validity (see Adcock and Collier 2001): Investigating whether we are indeed “measuring what we think we are measuring” (King, KEOHANE, and Verba 1994, 25) is nearly impossible if the argument is presented exclusively in operationalized form. Second, such a separation induces the author to specify which parts of the explanation are strictly historically—contextually specific and to differentiate them from the parts that constitute a “potentially generalizable model” (Sewell 1996, 270) in the sense of capturing insights that should, under certain conditions, be applicable more broadly. This differentiation, in turn, facilitates the identification of other cases, if any, on which the argument could subsequently be tested.5

5 For instance, in a paper on popular support for European Union membership in the 1994 Austrian referendum (Büthe, Copelovitch, and Phelan 2002), we start deductively from general theories of European integration and political economy to model support as a function of, among other factors, the economic interests of industrial sectors. We then work inductively from a qualitative analysis of the public debate before the referendum to specify those industrial sectors whose interests had salience for the public and might therefore be expected to have affected the outcome of the referendum. We test the resulting specified model quantitatively, using referendum results from Austria’s 121 districts as the dependent variable. The implied oversight of this caveat renders problematic the following methodological advice by Bates et al. (1998, 16): “The authors derive [testable hypotheses] from theory; but when the case materials do not confirm their expectations, they do not respond by rejecting their models. Rather, they respond by reformulating them and by altering the way in which they think about the problem.” This is not to say that historical works that do not use models cannot be very insightful. Barrington Moore’s (1996) classic, *Social Origins of Dictatorship and Democracy*, for instance, presents an argument that is so highly contextualized that Victoria Bonnell (1980, 170) finds that “his generalizations . . . cannot be reduced to . . . models.” And yet, it has provided insights and inspiration to several strands of research on regime change, class formation, and processes of social transformation. Similarly, the theoretical model remains largely implicit in Philip Nord’s (1995) *The Republican Moment.* Yet its rich and vivid account of the transformation of Parisian civil society from the 1860s to the 1880s, focusing on the emergence of democratic cultures and norms of conflict resolution in autonomous spheres of civil society prior to the successful political democratization of 1871, is surely very insightful for theorists and practitioners of democratization.
When success breeds success, when variables feed back into themselves, we have an exciting story to tell, but unless we know its metaphors [its model]... we have no way to tell it.
McCloskey (1991, 36)

MODELS AND HISTORY: THE PROBLEM OF TEMPORALITY

Several scholars have suggested in recent works that the social scientific study of “history” raises particular methodological problems (e.g., Bates et al. 1998; Elster 2000; Lustick 1996; Pierson 2000). Yet none of them defines what he means by history, not even Terrence McDonald (1996b), in his programmatic essay, “What We Talk about When We Talk about History.” So what is it that makes history distinctive?

Conceptualizing history as “what historians do” seems not very fruitful. Notwithstanding some real differences in stance and approach to evidence separating political scientists and political historians (Ingram 1997), what historians study is often studied with similar methods by political scientists, sociologists, economists, and others. Whether or not the study of history raises particular methodological problems for modeling or the presentation of empirical information would then become a matter of contested disciplinary conventions rather than of characteristics of the subject matter (Tilly 1990).

A conception of history as the study of “things of the past” is also not very promising. Studying past events may—but need not—require the researcher to work with sources that call for particular methodological, historiographic tools (McNeill 1998, 4f). Moreover, everything that we can study empirically must already have happened and, in that sense, is a “thing of the past.” If “history” is distinctive, then it must be conceptualized as a set of phenomena with distinct characteristics.

I submit that history—as an object of study that may require a distinct approach to theorizing and to the presentation of empirical information—must consist of macroprocesses that cover an extensive temporal space. How does this definition raise particular methodological issues? If the process itself is our explanandum, then isolating events from the historical process within which they occur risks depriving us of understanding because any one event in such a temporal sequence, far from being an independent observation, is meaningful only if seen as part of the larger process (Elias [1937] 1997, 80ff, [1939] 1997, 390; Mink [1966] 1987, 64f, 67, 80, 82f). Temporality thus becomes the defining characteristic of “historical” explananda. There are, to be sure, several ways of conceptualizing temporality (Aminzade 1992; Griffin 1992; Sewell 1996), but not least since time itself can be understood as a social construct; and the significance of the passing of time depends upon the level of abstraction at which we work.8

But whichever specific conception of temporality we adopt, it calls our attention to the issue of sequence, and it injects a dynamic element. Dynamics complicate the modeling task; sequence enables it.

Dynamics

The institutions within which actors interact are social constructs, as are the aggregate actors that populate so many of our models in political science. Due to factors such as uneven growth, increasing or diminishing marginal utility, and accumulation or ratcheting effects (e.g., Fearon 1998; Pierson 2000; Thelen 1999, 392ff) as well as the tendency of actors to attempt to manipulate or escape constraints (Almond and Genco 1977, 493, 518f), the passage of time makes it, ceteris paribus, more likely that institutions, actors themselves, and their preferences may change. Recognizing this dynamic aspect of temporality does not mean that everything is constantly in flux. In fact, institutions and aggregate actors can be extremely stable for a long period of time (Sewell 1996, 264). But the possibility of change implies that explanations of temporally large processes must allow for change in the constitution of actors (e.g., Ertman 1997; Spruyt 1994; Tilly 1998, 7ff) as well as for change in their preferences. In the sense of the inherent dynamism of temporality, then, history is “the study of changes of things that change” (Herbert Butterfield, as quoted in Schroeder 1997, 67). Models of “history” must explain stability rather than assume it.9

Because changes in preferences are all too easily invoked to explain changes in behavior and to claim the unsuitability of models that hold preferences constant, it is important to specify what I mean by preferences.10 I follow Jeffry Frieden in defining preferences as those interests of a given actor that determine how the actor rank-orders the possible outcomes (1999, 42). Preferences thus must be differentiated from the actions that the actor may undertake to achieve any particular outcome (“strategies” in game theory parlance), but they should also be sufficiently specific to the situation that they can unambiguously yield a rank-ordering of the outcomes that would result from the conceivable actions in that situation. Consequently, “what are considered preferences in one [context] might be strategies in another” (Frieden 1999, 41).

A change in preferences becomes more likely over time for two reasons. First, new ideas arise over time. While Mark Blyth (1997) rightly warns against a facile attribution of causal force to ideas, ideas can indeed

Footnotes:

8 The conception and increments of time that are relevant when the process we want to explain is a vast sweep of state formation (e.g., Hintze [1906] 1975; Tilly 1975, 1992) will surely differ from those that might be relevant when we are concerned with the decision-making process that leads to a particular policy decision.

9 This is one of the strengths of Barry Weingast’s chapter in Analytic Narratives: His model of U.S. federal institutions allows him to explain why policy remained stable over several decades despite major economic and demographic changes that significantly increased the numerical strength of groups opposed to the status quo and changed some of their preferences (Weingast 1998, esp. 161f, 184f, 188).

10 I thank one of the anonymous reviewers for thoughtful suggestions for clarifying this issue.

This content downloaded from 130.132.173.87 on Wed, 08 Jul 2015 14:13:49 UTC
All use subject to JSTOR Terms and Conditions
cause a change in preferences by making altogether new outcomes available, which forces a reordering of the outcomes. The rise of Keynesianism provides an illustration of this phenomenon. At the outset of the Great Depression, certain social actors failed to rank first among their preferences a Keynesian demand-side stimulus policy and failed to influence policymakers accordingly, even though with hindsight the adoption of such a policy would have been the outcome that maximized their material utility. Later, similarly situated actors indeed preferred such a policy over alternatives and employed various strategies to achieve the adoption of Keynesian policies by governments. This finding does not suggest that, at the earlier point in time, the actors were less instrumentally rational in ranking the outcomes and selecting their strategies accordingly—the “outcome” of Keynesianism just had not been formulated yet (Hall 1989).

Second, ideas and especially norms, being social constructs, can change (Ball, Farr, and Hanson 1989; Ruggie 1983; Wendt 1999, esp. 113ff). An example from the literature on regime change illustrates the resulting change in preferences nicely: Assume that policies are at least in part a function of regime type, and assume further that regime type has no effects on the citizens other than through policy (there are no normative commitments to a particular regime type). Social actors (individuals or groups) may, under these conditions, be expected to have preferences over (i.e., rank-order) the regime types based on which regime type affords increased relative to all other outcomes (e.g., Bermeo 1971; Hallgarten 1952; Lepsius 1978; Przeworski 1991, passim, esp. 51ff). If actors now develop over time a normative commitment to a democratic form of government, they will probably rank-order the possible outcomes of a regime change differently because the utility that they assign to the outcome “democracy” has increased relative to all other outcomes (e.g., Bermeo 1992; Di Palma 1990; Weingast 1997). The actors may still rank-order their strategies as before and may therefore choose the same action (for instance, acquiescence to the rule of the current nondemocratic regime if the perceived risks and costs associated with doing anything else render other possible strategies prohibitive), but their preferences have changed.11

The dynamic quality of temporality suggests that models based on assumptions of stable institutional contexts, stable preferences, and constant units for which we record variable, independent attributes at any given point in time would be unsuited if we are concerned with explaining history, understood as a macroprocess. Yet taking temporality seriously does not require abandoning modeling as such. In fact, standard game theoretic models can incorporate dynamic elements (for an introduction, see Brams 1994, and Gibbons 1992, 55ff, 173ff), and the extensive form used to depict and analyze such dynamic games indeed “take[s] sequence into account” (Bates et al. 1998, 14)—although it is based on a truncated conception of temporality in that, within such game models, “actual chronology is important only insofar as it influences what one player knows about the actions of the second” (Kreps 1990, 18). Models of historical processes, in contrast, need to derive the constitution of the actors and explanatory variables such as actors’ preferences within the model in order to allow for change (cf., e.g., Jackson and Nexo 1999, 302ff). At the same time, such endogenization of explanatory variables to capture dynamic change (and explain stasis where it occurs) does not require a fundamentally different approach to theorizing. Some recent work in economics, for instance, relies on only minor modifications of rational choice assumptions to develop dynamic models of preference formation (e.g., Becker 1996), which could be used as a building block of a larger model of an historical process. Evolutionary models developed in biology seem to be well suited to being adapted to explain the sociopolitical processes of persistence and change in the knowledge, values, and habits to which we customarily refer as “culture” (Boyd and Richerson 1985). In sum, endogenizing explanatory variables does not require a fundamentally new approach but can be achieved through building on, or adapting, various existing types of models.

Endogenizing explanatory variables, however, comes at the expense of parsimony or worse: Scholars who seek causal explanations usually frown upon endogenization because when the dependent variable is not only explained by, but also (partly) explains the independent variables, we run the risk of circular reasoning. Can we avoid this problem? Sequence provides the answer.

Sequence

Sequence allows us to endogenize the explanatory variables without having to abandon modeling and scientific aspirations because it enables us to avoid circular reasoning. Endogenization involves incorporating into the model some variation of causal feedback loops from the *explanandum* to the explanatory variables. In a static model, such feedback loops make the argument circular. Determining causality then becomes impossible. The sequential element of temporality, however, gets us around this problem, because it allows us to have causal feedback loops from the *explanandum* at one point in time to the explanatory variables at a later point in time only.12

The enabling effect of sequence is nicely illustrated by an example of a causal feedback loop in the chapter on education in Abram de Swaan’s (1988) *In Care of the State*, where the author seeks to explain the historical process by which elementary education, once seen

---

11 Higher-level preferences surely remain the same (in this case, for instance, the preference for longer life or safer possessions, which may be threatened by the current regime’s sanctions against pro-democracy activists). But—and this is the key to the differentiation between preferences and strategies used here—those unchanging higher-level preferences are insufficient by themselves to explain the rank-ordering of outcomes in the situation at hand.

12 Note that anticipated reactions can undermine the assumption that events at time *t* are independent of events at time *t + 1.
as an individual and local responsibility, came to be provided through nationwide collective and compulsory arrangements, administered or at least regulated by the state. In de Swaan’s model, an initial, partial success in increasing the scope and raising the quality of public education increases the opportunities for cross-regional commercial activities and the effectiveness of the state bureaucracies by raising the uniformity of language and knowledge. Over time, both of these consequences of increased and improved education should be expected to swell the ranks of the “metropolitan elites” (supraregionally trading entrepreneurs and central state bureaucrats) who favor the provision of elementary education as a public good. A change in the explanandum, education, thus leads to a change in the relative power of the actors in the conflict over the scope and quality of education at a later stage. It strengthens the proponents of widened and improved elementary education at the expense of the opponents, such as local elites and clergy, and in turn should lead to a further increase in the scope and quality of education.

Time itself thus becomes an element of the causal explanation, a factor in the model. But time does not function as a standard explanatory variable that directly affects the explanandum—otherwise, its effect could easily be expressed through a linear differential equation (McCloskey 1991, 22ff). Rather, it operates in the background to affect several explanatory variables in a variety of ways. Models that seek to help us explain historical processes qua processes therefore must explicitly incorporate a temporal dimension and consider carefully how each explanatory factor is affected by the passage of time in the process that we are trying to explain.

In sum, the importance of temporality is the distinctive characteristic of “historical” phenomena as objects of study, which raises particular methodological issues. The dynamic element of temporality complicates the modeling task by demanding that our models allow for the possibility of change (and hence explain rather than assume stability), but the incorporation of a sequential element enables us to do this without running the risk of the circular reasoning often associated with endogenizing the independent variables. Yet temporality raises additional problems when we move to empirical testing.

Model building is important for working out the internal logic of a chosen set of assumptions and relationships. But rigorous empirical analysis is needed to ensure the relevance of those assumptions and relationships.

Lazonick (1991, 303, emphasis added)

CONFRONTING MODELS WITH DATA: HISTORICAL NARRATIVES

Ways of confronting a model with data range from showing the consistency of the theoretical argument with one empirical observation of the explanandum (Eckstein’s [1975, 108ff] “plausibility probe”) to a test on a large sample that meets the criteria of statistical analysis (irrespective of whether we employ a quantitative or qualitative methodology). Where on this continuum a given work falls does not affect its usefulness or insightfulness but should affect our confidence in its conclusion, keeping in mind that any single work usually exists in the context of larger theoretical and empirical literatures. At any given point along the continuum, multiple analytical techniques are at our disposal. From the viewpoint of the presentation of empirical information, we can differentiate broadly between a primarily quantitative presentation, based on statistical analysis, and narratives, based on qualitative techniques such as process-tracing. In this section, I spell out what makes narratives a particularly suitable form of presenting empirical information when we want to test models about historical processes, and I suggest ways of structuring narratives to attain this benefit.

Why narratives? Practical constraints, such as too few instances of a given macrohistorical process, may inhibit the use of statistical techniques. Yet time series analysts have long used lagged variables, and in recent years some scholars have begun to develop more sophisticated procedures to control for time dependence in political phenomena (Beck and Katz 1996; Beck, Katz, and Tucker 1998). In short, we have very suitable statistical tools for the empirical analysis of models with dynamic and sequential elements, although if the nature of the explanandum indeed is such that all stages within each instance of the historical process (the explanandum) are interrelated, one would need to incorporate into the statistical model a lag for each variable at each stage of the process prior to the final stage, with a corresponding rapid decrease in the degrees of freedom.13 Moreover, various elements of the model and additional implications may be separately tested using statistical techniques. Statistical methods can thus be used in the empirical analysis of historical processes and may be particularly valuable when complementing and reinforcing insights gained from qualitative techniques (cf. Beck, Katz, and Tucker 1998; Berg-Schlosser and Quenter 1996; Mahoney 1999).

A preference for narratives, then, is due not to the unavailability of analytical techniques that lead to other forms of presenting our results, but to particular strengths of the narrative form. The most important of these strengths is that narratives, in addition to presenting information about correlations at every step of the causal process, can contextualize these steps in ways that make the entire process visible rather than leaving it fragmented into analytical stages. Moreover, narratives allow for the incorporation of nuanced detail and sensitivity to unique events, which may be necessary to understand the particular manifestation of an element of the model but which are beyond the model.
In de Swaan's narrative of the introduction of state-provided education in The Netherlands, for instance, he notes how the Napoleonic invasion made the central metropolitan elite suddenly exceptionally independent of local elites and clergy, thus strengthening them in the societal group conflict over the expansion of the state, including the expansion of public education. This unique event affected the speed of change and thus the particular manifestation of the process, but it did not change the general dynamic of the process as captured by the model.

At the same time, narratives must not revert to untheorized historical accounts, invoking extraneous factors in an ad hoc fashion, because such accounts are not useful as a test of the causal propositions. How can we avoid this problem? The model itself can help us write narratives that are useful as a test of the argument.

The theoretical model can be used to structure the narratives. As we know from Arthur Danto's (1965, 149ff) thought experiments, even the imagined "ideal chronicler" who records every action and event in perfect chronological order cannot provide a complete history, let alone a causal explanation. Any historical narrative therefore must simplify "reality" by designating some elements as salient and omitting many more as not significant (McClelland 1975, 75ff). The model can help by providing the criteria for what is salient: The actors identified in the model constitute the actors of the narrative, which traces their goals, beliefs, and actions. Within each narrative we thus employ process-tracing (Bennett 1999; George and McKeown 1985).

The influence of other elements of the model, such as how temporal progression affects the actors and their preferences, should be systematically described. Beyond the elements identified in the model, however, additional context-specific information should be minimized. Information that is extraneous to the model should be provided only insofar as it affects salient elements and is needed either to understand the relationship between these elements or to appreciate the contingencies of a particular historical process. If the narrative cannot be written in terms of the model, something is wrong with the model.

Using the model in this way to discipline the narrative ensures equivalence in the sense that each narrative contains the same (or at least functionally equivalent) elements. Each narrative thus becomes a unit or "observation"—a "plausibility probe" to test the causal argument, though we may not be able to assume independence for testing purposes (Sewell 1996, 258f). And most importantly, the model allows the analyst to overcome the problem of deciding what matters for the narrative (cf. Mink [1978] 1987, 187)—the problem that leads to the ad hoc-ness of many inductive historical explanations.

**HISTORICAL NARRATIVES AS DATA: FOUR PROBLEMS AND (PARTIAL) SOLUTIONS**

Notwithstanding the above strengths of narratives, their use as data brings to light four problems and limitations not identified in recent methodological work and debates on history, narratives, and social science theory. The first two problems are particular to the use of narratives as data in conjunction with models of historical processes; the third and fourth apply to any use of narratives as data.

The first problem concerns defining or delineating historical sequences as distinct, or what Bearman, Faris, and Moody call "casing" (1999; see also Mink [1966] 1987). If we conceive of history as a continuous stream of interrelated events, then at the logical extreme there is no beginning and no end. A narrative, however, inherently imposes a beginning and end onto the historical record. To be sure, all empirical work—whether large-N statistical or small-N case study work—needs to justify the boundaries of its units of analysis, especially when proceeding on the assumption of independence of observations (cf. King, Keohane, and Verba 1994, 222). But the emphasis on the interrelatedness of events across time makes this problem particularly acute for the empirical testing of models of history, while the literary and "aesthetic" (Topolsky 1998) qualities of the narrative tend to obscure it.

The model can provide a deductive, albeit only partial, solution to this problem. By specifying the *explanandum* in general terms and theorizing temporality explicitly, the model delineates a sequence as distinct—based on our research objectives—in the potentially infinite space of time. Note, however, that the *explanandum* itself provides a justification for choosing the starting and ending points of the narrative only insofar as the historical process we seek to explain can plausibly be said to have had a clear starting point (e.g., an exogenous shock) and to have "run its course." A number of sociologists have in recent years developed alternative *inductive* procedures, often grounded in the boundary specification approach of network theory, for delineating an event sequence as a distinct process and hence justifying a distinct narrative with clear end points: "event structure analysis" (Heise 1993), "abstraction and generalization of interactions" (Abell 1993), "interaction process analysis" (Kosaka 1993), and "bicomponent analysis" of event populations (Bearman, Faris, and Moody 1999). However, questions remain about the logic underpinning these procedures (see Abell et al. 1993), and I doubt their general applicability to macrosocial phenomena. This leaves us with a clear specification of the *explanandum* as the only generalizable, methodological justification.

---

14 Carefully tracing the positions taken by each of the actors identified in his model, de Swaan shows, for instance, that over time (for reasons consistent with the model, such as a growing cohesion of a group-social actor), the demands of working-class parents intensified, whereas "time" reversed the preferences of industrialists.

15 I address the issue of independence below, when discussing the use of multiple narratives as a remedy for what I identify as the "third problem" of historical narratives as data.
for making choices about where to begin and end a narrative.

De Swaan (1988), whose use of historical narratives is exemplary, achieves such a delineation of a temporal sequence by defining his explanatory as the process by which elementary education was transformed from a private and local affair into a responsibility of the state, provided at a basic level throughout the territory within its reach, according to certain standards (e.g., a uniform language) and financed through compulsory measures. This definition of the explanatory suggests that the narrative must start with the initial moves (ideational or practical) away from the previous local and private system of schooling but need not be concerned with that system's prior history. And the narrative can end when uniform, state-financed elementary education has been established in the country in question, without needing to concern itself with the subsequent evolution of the educational system or other related aspects of the welfare state. The specification of the explanatory thus provides the criteria for choosing the beginning and end of the narrative.

The second problem concerns the need to "conclude" the narrative while the process may be ongoing, which restricts the "generality" of our conclusions (King, Keohane, and Verba 1994, 137). If we use feedback loops in our models, through which a change in the value of the dependent variable is hypothesized to change the value of some or all of the explanatory variables at a later point in time, time itself acquires an explanatory role (in interaction with the original explanatory variables). As long as only the explanatory variables, and not the explanatory itself, are affected by time in this way, truncating time by setting an end point for the narrative does not introduce the selection bias that is caused by truncating the range of the dependent variable (King, Keohane, and Verba 1994, 128ff). But more than with conventional independent variables, of which we may restrict the range, we have to be very careful about assuming consistent continuity of the relationship between the explanatory variables and the explanatory beyond the investigated period. When time itself becomes a factor in the model, as discussed above, we have to consider carefully how each explanatory factor is affected by the progression of the process that we are trying to explain. The extent to which we can expect the effect of time on the explanatory (through the other independent variables) to continue as observed during the time period covered by the narrative depends on the tenability, beyond this time period, of the assumptions that we had to make to model the effect of time. This limitation of historical narratives affects the confidence we can have in the generalizability of the insights we gain from them. The solution, however, lies not in modestly claiming that our conclusions cannot be generalized but, rather, in paying careful attention to temporality in both the model and the narratives and specifying the implications for generalization accordingly.\footnote{The applicability of a historical model by no means needs to end with the time period on which it was empirically tested. Otto Hintze (1906) 1975, for instance, long ago provided a fascinatingly simple example of what we would today call a "second image reversed" (Gourevitch 1978) model of state formation, capturing an essential element of a causal mechanism that might explain not only the various transformations of the "organization of the state" over the thousand-year stretch of European history to which he applies it, but also phenomena that occurred many decades after the formulation of his model, such as European integration and the later parts of the "third wave of democratization" (Huntington 1991, esp. 85ff).}

Third, what is the status of any narrative's truth claim? As Mink ([1978] 1987, 199) put this "dilemma of the historical narrative":

\textit{...As historical it claims to represent, through its form, part of the real complexity of the past, but as narrative it is a product of imaginative construction, which cannot defend its claim to truth by any accepted procedure of argument or authentication.}

The problem here, as Andrew Norman (1991, 131) points out, is not that "a discursive representation has a structure that that which it represents does not." There is no necessary link between discursive structure and misrepresentation. Moreover, some objective criteria for assessing a narrative's truth claim exist, such as the extent to which an author is able to provide from the historical record evidence that, after having been subject to standard historiographic procedures, supports the author's "story." It is therefore hardly necessary to equate historians with novelists or fiction writers (Gaddis 1992/93, 56). Rather, the problem is that, because facts never speak for themselves (Lustick 1996), there is an interpretative element that cannot be evaluated from within the narrative without circular reasoning. To assess this aspect of the narrative, the reader will have to draw on knowledge from sources external to the narrative at hand; i.e., the reader will need to know something about the historical period in question from other sources—sources that themselves are bound to have narrative qualities. Consequently, it is more meaningful to endorse good narrative work as "plausible," "persuasive," or "compelling"—as seems to be the practice among historians—rather than "true" or "right," though we certainly may find some narrative work that is poor and even plain "wrong," such as when its interpretation is marred by logical inconsistencies or makes incorrect assertions about the chronology of events.

Moreover, that the truth claim of a narrative cannot be assessed from within the narrative itself also has important implications for the utility of narratives as data on which to test hypotheses derived from models. Presumably, no one will consciously publish a model with empirical information that directly contradicts it. Except for discarded alternative explanations (see below) and possible oversights, then, the vulnerability of the model to empirical disconfirmation is ultimately also external to the narrative.

However, the use of multiple narratives may increase our confidence that the model indeed captures the key dynamics of the process. Using multiple narratives is appropriate because models should be applicable to more than a single instance if they have the benefits...
of capturing what is generalizable. Having assured that each narrative contains the same elements, we can employ each narrative as a unit or “observation” to test the causal argument. To be sure, we are unlikely to be able to furnish sufficiently many historical narratives to perform statistical tests, and by virtue of treating each historical narrative as a self-contained unit so as to incorporate temporal progression into each “observation,” we have to reject any attempt to make “many observations from few” (King, Keohane, and Verba 1994, 217ff). Moreover, there is the problem of independence.\(^{18}\) As William Sewell and others point out, if separate historical instances of a process are treated like ideal–typical scientific experiments, such treatment assumes that they are independent of each other. This assumption is problematic because, unless there is a perfect informational separation between the instances, actors in one (later or perhaps contemporaneous) instance will have knowledge of the constitutive actions and the outcome of the other instance(s) (Sewell 1996, 258ff). Such knowledge will violate the assumption of independence—for purposes of testing a model—only if it changes elements of the process as modeled (e.g., the preferences of actors), which is by no means a necessary, though a possible and indeed quite likely consequence of having knowledge of prior or other instances.\(^{19}\) Where empirical analysis suggests such an effect, we therefore cannot rely upon the statistical logic of the traditional comparative method (Lijphart 1971). But as “plausibility probes” multiple narratives are very useful. Increasing the number of confirming narratives does not in any way “prove” the model (Mohr 1996, 118ff), but in light of the temptation of inductivist modifications of a given model, the ability of a model to withstand the difficult test of application to different occurrences of the explanandum without ad hoc alterations makes more plausible that it has captured the central, generalizable dynamics rather than unique elements of a particular case.

In fact, it is the use of multiple narratives that makes Theda Skocpol’s (1979) States and Social Revolutions, Margaret Levi’s (1998) “Conscription: The Price of Citizenship” (1998), and, especially, de Swaan’s (1998, 52–117) “The Elementary Curriculum as a National Communication Code” so convincing. To be sure, the extensive methodological debate about Skocpol’s book has produced “little consensus” about the status of her narratives (Mahoney 1999, 1156), and her consciously inductive approach makes it questionable to treat her narratives as data to test a model, but the book derives most of its persuasive power precisely from her ability to narrate three instances of social revolutions in the same terms. Multiple narratives work even more effectively in Levi’s “Conscription.” Though it is sometimes difficult to see how the various parts of her complex model structure the three narratives, each of her three accounts of the abolition of “buying one’s way out of military service” (Levi 1998, 109), in France, the United States, and Prussia, shows the same, separately theorized process at work. Finally, de Swaan uses five separate narratives (for the United States, France, Britain, Germany, and The Netherlands) to test his—separately theorized though not fully explicit—model of the historical process by which elementary education becomes the domain of the state. In all of these works, what persuades the reader of the validity and usefulness of the model is that for each of the countries, as Sewell (1996, 262) puts it with respect to Skocpol, the specific historical process “can be narrated convincingly in terms of the operation of analogous causal processes, which...[thus] make sense of numerous details that otherwise would seem purely accidental.” In this sense, providing multiple “plausibility probes,” far from being “fruitless[ly] repetitious” (Skocpol and Somers 1980, 191), should enhance confidence in the explanation. This way of combining models with narratives allows us to provide a “scientific” causal explanation of historical processes, without depriving them of their process character.

Fourth, due to the limited truth claims of narratives, those who use historical narratives as empirical evidence for a causal explanation will probably fail to assess alternative explanations and, if they try, will fail to convince skeptics. Trying to assess alternative explanations would entail providing a second set of narratives, similarly structured by the underlying model of an alternative explanation for the same phenomenon. To be sure, such alternative narratives could conceivably be supplied by the same author just as those working quantitatively often operationalize and test alternative explanations. In fact, Graham Allison (1971) employed three separate narratives to test three alternative, competing models of foreign policymaking in Essence of Decision.\(^{20}\) But alternative narratives by the same author serve primarily as a rhetorical device in support of the primary, favored explanation. The interpretative freedom of the author makes it unlikely that less convincing alternative narratives would be accepted as sound evidence of the failure of the alternative explanations. This, however, is not a serious problem. As Morris Fiorina notes, meaningful alternative explanations are much more likely to be advanced by others whose “perspectives and commitments” allow them to argue as strongly as possible in support of those alternatives. All such explanations must then be subjected to the collective assessment of the scholarly community at large (Fiorina 1995, 92).

**CONCLUSION**

I have pursued two objectives: (1) I have sought to clarify, based on a broad conception of modeling and an explicit definition of “history,” the difficulties—and possibilities—of modeling historical processes; and

\(^{18}\) I thank one of the anonymous reviewers for thoughtful comments on this issue.

\(^{19}\) Statisticians therefore speak of seeking “independence of the error terms,” which is a purely empirical matter, rather than “independence of observations.”

\(^{20}\) I thank Steve Solnick for reminding me of Allison’s use of alternative narratives.
(2) I have sought to make the case for narratives, identify four overlooked problems with their use as data to test models about historical processes, and offer at least partial solutions to these problems.

I have argued that modeling is rendered more difficult when dealing with “history” due to the importance of the dynamic and sequential aspects of temporality for processes that can be methodologically meaningfully distinguished as “historical.” But social scientists interested in the study of such “history” need not return to narratives that, detached from any explicit causal model of the historical process they seek to capture, implicitly make—often sweeping—causal claims based on hidden assumptions and unspecified causal mechanisms. There is no inherent incompatibility between the deductive modeling of causal relationships and the study of history. Quite to the contrary, for social scientists—and historians—who seek causal explanations of historical processes, making their models explicit and subjecting them to separate theoretical examination have several advantages. Most importantly, the greater transparency of a model (compared to a purely “operationalized” argument) facilitates the scrutiny of the argument’s assumptions and internal logic, which is particularly desirable for inherently complex historical phenomena. Further, models help clarify what insights the author suggests to the reader, by inducing the author to specify which elements of the argument are, in the author’s mind, uniquely specific to the historical setting and which parts of the argument are potentially generalizable. In fact, the model should be epistemologically prior to, and independent of, the narrative(s) used to “test” it, in order to enable the model—rather than the narrative—to do the explanatory work (cf. Somers 1996, 79ff).

Regarding the benefits of using narratives, I have emphasized the ability of narratives to contextualize any given stage in the process, which is essential when we seek to capture empirically the process for which the model stipulates an explanation. To be useful as a test of a deductively sound model, a narrative should be structured by the model in that the presentation of empirical information follows the model’s identification of the historical setting and which parts of the argument are potentially generalizable. In fact, the model should be epistemologically prior to, and independent of, the narrative(s) used to “test” it, in order to enable the model—rather than the narrative—to do the explanatory work (cf. Somers 1996, 79ff).

At the same time, using narratives as data to test a model that seeks to capture an entire historical process raises four problems not identified in recent debates. I have specified these problems and analyzed their implications for the usefulness of narratives as data. I have suggested ways of solving or minimizing the problems, including the use of multiple narratives to increase our confidence that the model indeed captures the key elements of a causal explanation of the processes that we are seeking to explain.

So does the study of history, understood as the analysis of processes that span a large temporal space, require a distinct approach to theorizing and to the presentation of empirical information? I conclude that it indeed engenders particular problems and requires us to be attentive to issues that may not arise in analyses of phenomena where temporality plays no role. But the study of history does not require a fundamentally different approach. Notwithstanding some remaining limitations, historical narratives can provide strong support for the plausibility of a historical model, maintaining a basic commitment to a scientific approach while overcoming several of the problems of the compatibility of historical narrative and social science identified by critics of the “protoscientific” ambitions of historical analyses. And it may be that at the level of broad historical phenomena, where the temporal dimension is crucially important to any explanation, providing such support for the plausibility of the argument is the best we can hope for, at least from any one scholar. Science is, after all, a collective enterprise (Weber [1919] 1946).

REFERENCES


Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies

HERBERT P. KITSCHELT*

Since the 1960s, successive protest movements have challenged public policies, established modes of political participation and socio-economic institutions in advanced industrial democracies. Social scientists have responded by conducting case studies of such movements. Comparative analyses, particularly cross-national comparisons of social movements, however, remain rare, although opportunities abound to observe movements with similar objectives or forms of mobilization in diverse settings.

A social movement that lends itself to cross-national study is the anti-nuclear power movement, which swept across the political landscapes of America and Europe in the 1970s. In some countries, the nuclear power conflict reached an intensity unprecedented in the history of technology controversies. So far, the opportunity for a theoretically-orientated and controlled comparison of anti-nuclear movements has not been seized, for while case studies of nuclear power conflicts generate a wealth of descriptive detail, individually they are not suited to the task of arriving at a generalized understanding of the factors that determine the dynamics of social movements.1

This article is an attempt to use some of the rich detail of the existing case studies to construct a systematic comparison of the anti-nuclear power movements in France, Sweden, the United States and West Germany. All four countries have experienced intense conflicts over nuclear technology, but anti-nuclear movements in each have pursued a different strategy and have had a different impact on overall energy policy. I shall argue that a particular set of variables is most useful for explaining these variations,

* Department of Political Science, Duke University, Durham, N.C. For helpful comments on an earlier version of this paper, I would like to thank Peter Lange, Anthony King, Sidney Tarrow, and two anonymous reviewers for the Journal.

1 The non-theoretical literature includes several useful handbooks, written by anti-nuclear activists, about the development of nuclear power conflicts in a number of countries, as well as several descriptively rich comparative analyses written by academic observers. Representative handbooks include: Projektbereich Ökologie der Vereinigten deutschen Studentenschaft, Bochum, Atomenergie International: Atomprogramme und Wilderstand in 28 Ländern (Bochum: Druckladen, 1978); Anna Gyorgy and friends, No Nukes: Everyone’s Guide to Nuclear Power (Boston: South End Press, 1979); and Lutz Mez, ed., Der Atomkonflikt (West Berlin: Olle und Wolters, 1979). Representative academic analyses include: John Surrey and Charlotte Huggett, ‘Opposition to Nuclear Power: A Review of International Experience’, Energy Policy, iv (1976), 286–307; Dorothy Nelkin and Michael Pollak, ‘The Politics of Participation and the Nuclear Debate in Sweden, the Netherlands, and Austria’, Public Policy, xxv (1977), 333–57; also ‘The Political Parties and the Nuclear Energy Debate in France and Germany’, Comparative Politics, xii (1980), 127–41; and also The Atom Besieged: Extraparliamentary Dissent in France...
namely, a nation’s political opportunity structure. Political opportunity structures are comprised of specific configurations of resources, institutional arrangements and historical precedents for social mobilization, which facilitate the development of protest movements in some instances and constrain them in others. While they do not determine the course of social movements completely, careful comparisons among them can explain a good deal about the variations among social movements with similar demands in different settings, if other determinants are held constant. Comparison can show that political opportunity structures influence the choice of protest strategies and the impact of social movements on their environments. The latter, in particular, is a topic that has received little attention until recently.2

The explanation of the strategies and impacts of social movements suggested in this article differs from – but is not necessarily inconsistent with – those advanced by three other theoretical approaches: Marxist-macrosociological, microsociological and resource mobilization. Essentially, what distinguishes the approach taken here is the importance assigned to explaining movement variations, both in terms of mobilization and impact. Marxist-macrosociological analysis, for example, links the emergence of social movements to various stages in the development of socio-economic modes of production; and those following this approach have viewed the anti-nuclear movement as a member of a larger class of ‘new social movements’ that has been spawned by the systems of bureaucratic and technological control that regulate social life in late capitalism.3 What proponents of this approach do not explain is why the various national anti-nuclear protests have had such dissimilar careers,


in terms of both differential articulation and impact, in otherwise similarly constituted capitalist societies.

In sensitivity to the importance of explaining movement variations also characterizes microsociological approaches, which seek to explain the mobilization of protest and its impact on policy and institutions as direct consequences of the number and intensity of social 'strains' and 'grievances' or of the relative deprivation experienced by particular social groups. As has often been noted, strains or deprivations in and of themselves seldom explain variations in the dynamics of social movements. This is certainly true for the cases under study; each country's energy program presented its citizens with similar levels of grievances, but the national movements that emerged developed in distinct ways. As will be shown, political opportunity structures functioned as "filters" between the mobilization of the movement and its choice of strategies and its capacity to change the social environment. At most, we can say that the existence of strain and relative deprivation is a necessary but not a sufficient condition of social protest.

The explanatory approach suggested here is loosely linked to the relatively recently elaborated resource-mobilization perspective in social protest research, which conceives of social movements as collective and rational decision-makers that mobilize their followers and promote their causes with the best available strategies given limited cognitive and material resources. Most of the empirical studies that adopt this perspective, however, concentrate on those internal variables of movement mobilization that are deemed to be within an incipient movement's discretion, e.g., incentive structure in membership recruitment, internal organization, specification of goals and skills in forming coalitions with allies. In contrast, the emphasis of the present comparison of anti-nuclear movements is on relating the strategic choices

---


and societal impacts of movements to specific properties of the external political opportunity structures that movements face. Such institutional constraints have often been simply assumed, rather than systematically and comparatively used to explain social movements’ trajectories.7

A comparison of anti-nuclear protest movements in France, Sweden, the United States and West Germany is well-suited to discovering the effects of institutional constraints on social movement mobilization for several reasons. First of all, these four anti-nuclear movements share similar operational objectives, namely, to prevent the completion of nuclear power plants under construction, to prevent work from beginning on planned projects and, ultimately, to shut down existing nuclear facilities. Secondly, in all of the cases, nuclear power conflicts grew from localized, segmented conflicts about specific power plants into national movements and controversies in the same time period, soon after the first energy crisis of 1973–74.8 (Anti-nuclear movements

7 McCarthy and Zald, ‘Resource Mobilization and Social Movements’, p. 1236, for instance simply state that they have assumed the ‘modern American context’ for their theory. But the institutional context is, as Piven and Cloward, pp. 15–37, point out, an important determinant of movement mobilization that may vary. A greater emphasis on external political opportunity structures is found in some of the recent social movement research. See McAdam, Political Process and the Development of Black Insurgency; Tarrow, Social Movements; and Tilly, From Mobilization to Revolution, Chap. 4.

are treated here as complex aggregations of protest events at the level of entire countries, not as sequences of separable protest episodes at a more disaggregated level. Thirdly, the objective ‘threat’ of nuclear power was about the same in each country in that all governments were firmly committed to nuclear programs of approximately the same size and growth rates at the time that anti-nuclear protest became a national phenomenon. Each country, for example, expected to install one to two gigawatts of nuclear electricity generation capacity per million inhabitants by the late 1980s. Finally, as we shall see, the subjective sense of deprivation and grievance was quite similar. This assessment is supported by data about the social base of the movements indicating that the primary recruits were professionals and (public) service sector employees, farmers and property owners in the vicinity of proposed nuclear sites, students and young radicals, making each national movement an expression of ‘middle-class radicalism’. They shared not only similar social bases but similar opponents: each faced a pro-nuclear coalition comprised of nuclear scientists, engineering firms, utilities and promotional or regulatory state agencies.

EXPLAINING STRATEGIES AND IMPACTS OF SOCIAL MOVEMENTS

Political opportunity structures can further or restrain the capacity of social movements to engage in protest activity in at least three different ways. Firstly, mobilization depends upon the coercive, normative, remunerative and informational resources that an incipient movement can extract from its setting and can employ in its protest. In Western democracies, non-violent resources are crucial for the emergence of protest. Thus, if movements can appeal to widely shared norms, collect adequate information about the nature of the grievance against which they protest and raise the money to disseminate their ideas and information, the chances of a broad mobilization increase. Secondly, the access of social movements to the public sphere and political decision-making is also governed by institutional rules, such as those reinforcing patterns of interaction between government and interest groups, and electoral laws. These rules allow for, register, respond to and even shape the demands

9 Reliable quantitative data about the social background of anti-nuclear activists are hard to come by. But the case studies referred to in fn. 4 consistently identify these three groups of activists.

10 The alliances of pro-nuclear interests were very similar in all four countries during the early stages of the nuclear power debate. However, the reasons why these clusters of industrial and administrative interests are the logical outcome of nuclear technology development in the countries compared here are discussed in Herbert Kitschelt, ‘Structures and Sequences of Nuclear Energy Policy-Making: Suggestions for a Comparative Perspective’, Political Power and Social Theory, iii (1982), 271–308.

11 The concept of political opportunity structure is used here in a broader sense than that conveyed by ‘state structure’, a concept that has been used, and criticized, in recent discussions in the field of comparative public policy. See John Zysman, Governments, Markets and Growth (Ithaca, NY: Cornell University Press, 1983), pp. 291–300 and 347–9. Opportunity structure
of social movements that are not (yet) accepted political actors. They also facilitate or impede the institutionalization of new groups and claims. Thirdly, a social movement faces opportunities to mobilize protest that change over time with the appearance and disappearance of other social movements. The mobilization of one movement, for example, may have a 'demonstration effect' on other incipient movements, encouraging them to follow suit. And the simultaneous appearance of several movements contesting the institutions of social control often presents the best opportunity to maintain movement momentum and to change established policies.\(^\text{12}\)

In the four countries compared here, the temporal opportunity structures encountered by the anti-nuclear movements were quite similar; the protests reached a peak in the second half of the 1970s and they grew out of the broader environmental movement. Crucial differences, however, characterize the resource and institutional opportunity structures they faced.\(^\text{13}\) These configurations, which are relatively inert over time, may also be labelled as the 'political regimes' prevailing in each country. While they are not immutable, they respond only slowly to new policy demands. And inasmuch as they pattern policy demands and options independently of the preferences of shifting coalitions of interested political actors and social forces, they injection a decidedly non-pluralistic element into the policy formation process.

Students of social movements at times distinguish relatively 'open' political opportunity structures from relatively 'closed' ones and note that the dominance of one type or the other sets limits to the responsiveness that movements can expect.\(^\text{14}\) A particularly useful outgrowth of this research is the identification of a curvilinear relationship between openness and movement mobilization, which shows that very closed regimes repress social movements, that very open and responsive ones assimilate them, and that moderately repressive ones allow for their broad articulation but do not accede readily to their demands.

This conceptualization of opportunity structures is useful but somewhat one-sided, for it considers only the input processes of political decision cycles. The other side of the coin is that the capacity of political systems to convert demands into public policy also affects social movement mobilization and encompasses the concept of 'dominant policy style'. The latter is developed for a number of countries in Jeremy Richardson, ed., \textit{Policy Styles in Western Europe} (London: George Allen & Unwin, 1982). A recent analysis in this vein of labour movements is found in Peter Lange, George Ross and Maurizio Vannizelli, \textit{Unions, Change and Crisis: French and Italian Union Strategy and the Political Economy, 1945–1980} (London: Allen & Unwin, 1982).

\(^\text{12}\) The concepts of movement cycles and reform cycles are developed in Tarrow, \textit{Social Movements}, pp. 35–46.

\(^\text{13}\) Structures are those processes in a system that change at a rate so slow as to be fixed for the study of events that transpire over a short period of time. See Karl Deutsch, 'The Crisis of the State', \textit{Government and Opposition}, \textit{xvi} (1981), 331–41, at p. 332.

\(^\text{14}\) The distinction drawn between open and closed opportunity structures is used in Peter K. Eisinger, 'The Conditions of Protest Behavior in American Cities', \textit{American Political Science Review}, \textit{lxvii} (1973), 11–28.
impact: the output phase of the policy cycle also shapes social movements and offers them points of access and inclusion in policy-making.\textsuperscript{15} Indeed, this conclusion is supported by the many case studies which show that policies are often entirely renegotiated as they are implemented.\textsuperscript{16} Thus, the capacity of political opportunity structures to implement policies – as well as their openness to societal demands – ought to be seen to determine the overall responsiveness of politics to social movements.

While it is certainly the case that political opportunity structures vary among policy arenas within the same political regime, system-wide political properties and national ‘policy styles’ also play key roles in determining the dynamics of social movements. The nature of these properties and styles are of crucial importance because representatives of entirely new demands often cannot participate effectively in highly differentiated policy arenas and instead must appeal to actors and institutions in politics, such as parties, parliaments and courts, whose authority and decision procedures at least partially transcend those of particular policy arenas.

In this respect, at least four factors determine the openness of political regimes to new demands on the input side. (1) The number of political parties, factions, and groups that effectively articulate different demands in electoral politics influences openness. The larger this number, the more ‘centrifugal’ a political system tends to be and the more difficult it is to confine electoral interest articulation to the ‘cartel’ of entrenched interests that is represented by the established, bureaucratized parties. (2) Openness increases with the capacity of legislatures to develop and control policies independently of the executive. This is the case because a legislature is by definition an electorally accountable agent and is therefore much more sensitive to public demands, whereas only the uppermost positions in the executive are subject to such direct public pressure. (3) Patterns of intermediation between interest groups and the executive branch are another element shaping political openness. Where ‘pluralist’ and fluid links are dominant, access for new interests to the centres of political decision-making is facilitated. (4) Finally, political openness not only requires opportunities for the articulation of new demands, but new demands must actually find their way into the processes of forming policy compromises and consensus. For this to occur, there must be mechanisms that aggregate demands. Openness is constrained when there are no viable procedures to build effective policy coalitions.

In a similar vein, three operational dimensions characterize the capacity of political systems to implement policies. (1) National policies are implemented more effectively when the state apparatus is centralized. A complicated division of jurisdiction between a multitude of semi-independent government agencies and a federal stratification of state authority tends to

\textsuperscript{15} This term is used in Judith May and Aaron Wildavsky, eds, The Policy Cycle (Beverly Hills, CA: Sage, 1978) to describe public policy processes in terms of steps and stages.

\textsuperscript{16} This point is frequently stressed in implementation research. See Eugene Bardach, The Implementation Game (Cambridge, MA: MIT Press, 1977).
make policy implementation more cumbersome. (2) Simultaneously, government control over market participants is a key variable for government effectiveness in many policy areas. The degree of state control over the finance sector, the relative size of the public sector’s share of GNP and its share of total employment, and the state’s co-ordination, control or exclusion of economic interest groups in policy-making, are some of the factors that influence policy effectiveness. The greater is the control of economic resources and decision centres through political institutions, the more limited are the resources available with which to challenge policies. (3) Policy effectiveness is also determined by the relative independence and authority the judiciary enjoys in the resolution of political conflict. Policy implementation becomes more hazardous and cumbersome if courts are forums of political arbitration removed from executive branch control.

Differences in the openness and capacity of political regimes are continuous rather than discrete variables. Given the number of variables on each dimension, many combinations of openness and implementation capacity may occur. Nevertheless, for comparative purposes, one may roughly dichotomize each of the political input and capacity variables. Doing so shows that each of the four countries included in the present study represents a different configuration of regime properties (see Table 1). Space constraints rule out a detailed defence of this classification of opportunity structures, let alone an analysis of each structure’s historical origins or regime changes. Some classifications, however, are likely to be contested and therefore deserve a brief discussion.

<table>
<thead>
<tr>
<th>Political output structures</th>
<th>Political input structures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Open</td>
</tr>
<tr>
<td>Strong</td>
<td>Sweden</td>
</tr>
<tr>
<td>Weak</td>
<td>France</td>
</tr>
<tr>
<td>United States</td>
<td>West Germany</td>
</tr>
</tbody>
</table>

With respect to France, there is broad agreement on the effectiveness, though not the efficiency, of national policy-making. More debatable is the characterization here of the French political system as closed. There are,

17 This variable permits only a restricted, though important, scope of generalization across policy areas. While it is important as a determinant of most economic and social policies, there are obviously other policy areas where it does not come into play as a determinant of policy formation.

18 The divergent features of political regimes found among advanced industrial democracies can be traced back to the circumstances surrounding their state-building, their location in the world economy, the timing and speed of their industrialization, and the formation of class and group coalitions promoting specific regime forms. For the purposes of this article, however, the varying outcomes of political development in the four countries are taken as givens.
however, several features of policy-making that support this characterization: the executive branch is clearly dominant over a weak legislature and there is policy-making access for only a select number of interest groups. Moreover, the party system of the Fifth Republic exhibits centripetal tendencies. Increasingly, this has meant that two blocs, organized along the fundamental socio-economic cleavages of French society, vie for political power. Thus, the two main competitors in the party system have had difficulty in accommodating the demands generated by the cross-cutting cleavages of the ‘new politics’.19

Sweden’s political capacity may not be as high as that of France, but its unitary public administration, weak political judiciary and fairly high degree of control and concertation of the economy justify its characterization as a ‘high effectiveness’ polity, as compared with either West Germany or the United States. More questions, though, can be raised about the openness of Swedish politics. While societal corporatism may indicate the opposite, Sweden’s relatively differentiated, fractionalized party system and its consensus-orientated, responsive bureaucracy are all factors that weigh in favour of characterizing it as open. This is further supported by the fact that hitherto unrepresented new-politics issues, such as demands for participatory democracy, rights for students, the emancipation of women, comprehensive aid to less developed countries and, more recently, civil and socio-economic rights for immigrants have all been attentively registered by the political parties and have triggered policy innovations.20 In both respects, Sweden’s political process displays relatively great openness.21

---


America's political input structures exhibit fairly great openness to interest articulation but far less openness with respect to the aggregation of new demands. The comparatively strong position of the Congress, the lack of tightly integrated political parties, the relative openness of a deeply fragmented administration, all testify to the openness of politics in the United States. But the lack of structured systems of intermediation between interest groups, legislators and the political bureaucracy impose severe constraints on the capacity for political aggregation and innovation. New demands often 'evaporate' in the pluralist process of coalition formation or later on when a weak state agency has to renegotiate a policy with organized interests. The executive branch is territorially and substantively fragmented, has little control over the economy and must face an autonomous judiciary.  

The description of *West Germany* as a polity with a 'weak' capacity breaks with the efficiency myth with which German politics has often been falsely associated. The jurisdictional and territorial fragmentation of the state is great, the judiciary is quite autonomous, and the state is restricted with respect to both the choice of instruments and the resources at its disposal in the control of private market actors. At the same time, its centripetal party system, organized along class and religious cleavages, weak legislature and inaccessible executive make West German political input structures appear more like those of the closed French system than those of the more open American and Swedish polities.

How do these different national political opportunity structures affect the strategies and impacts of social movements? Two major hypotheses guide the present comparison of anti-nuclear movements. Firstly, with respect to strategies, political opportunity structures set the range of likely protest activities. For instance, when political systems are open and weak, they invite *assimilative* strategies; movements attempt to work through established institutions because political opportunity structures offer multiple points of access. In contrast, when political systems are closed and have considerable capacities to ward off threats to the implementation of policies, movements are likely to adopt *confrontational*, disruptive strategies orchestrated outside established policy channels.

Secondly, political opportunity structures facilitate or impede movement impacts, among which we may distinguish three types: procedural, substantive

---


and structural. Procedural impacts or gains open new channels of participation to protest actors and involve their recognition as legitimate representatives of demands. Substantive gains are changes of policy in response to protest. And structural impacts indicate a transformation of the political opportunity structures themselves as a consequence of social movement activity.

To elaborate further, the second hypothesis leads us to expect procedural gains to covary with the openness of political systems. Thus, open regimes should be more willing to accept new groups, as it is likely that at least some established political actors will seek to strengthen their own positions by allying themselves with the newcomers. This incentive is missing in closed systems, where policy-making is the prerogative of a circumscribed cartel of political actors. For substantive gains to be made, a polity must have not only relatively open institutions and policy-making procedures but a high capacity to implement policies. The more openness and capacity converge, the greater the likelihood of policy innovation. A variation should occur when a regime is closed and strong. In this instance, movement activities may prompt a limited range of elite-initiated reforms. Substantive gains are least likely to be found in weak regimes, be they open or closed. Here the likely outcome of protest activity is political stalemate, a situation in which neither old nor new policies can be implemented successfully. Finally, structural impacts will figure when a political system cannot bring about either procedural or substantive reforms. In this instance, a social movement will try to broaden its demands to include those for altering the existing political system fundamentally.

According to the logic of these hypotheses, social movements in the four cases under consideration ought to pursue distinct and different strategies and to have different policy impacts. These expectations are summarized in Table 2. Bearing them in mind, we turn now to the empirical data about the careers of anti-nuclear movements in France, Sweden, the United States and West Germany.

**Strategies of Anti-nuclear Movements**

What strategies have anti-nuclear protesters adopted? Assimilative strategies have included lobbying, petitioning government bodies, influencing public policy through referendum campaigns and partisan involvement in electoral contests. Additionally, movements have tried to affect policy implementation by participating in licensing procedures and litigation. Confrontational strategies have included public demonstrations and acts of civil disobedience, exemplified by occupations of nuclear plant sites and access roads. But the national movements have not chosen equally from among these two types of protest. Moreover, as Table 3 shows, the choice of strategy does not vary at random. Rather, it varies with the specific type of political opportunity structure.

---

24 Gamson, *The Strategy of Protest*, Chap. 3, introduced the important distinction between procedural and substantive impacts or ‘gains’ of movements, but omitted structural impacts on the political regimes themselves.
In the political systems where open decision-making processes prevail, anti-nuclear activists attempted to influence, directly and indirectly, legislatures and elections early on in the policy conflict. Thus, in the United States, environmental interest groups adopted an assimilative strategy that included, at the federal level, lobbying in Congressional committees and attempting to influence key sets of actors in the executive branch such as the Council on Environmental Quality, the Environmental Protection Agency and, to a lesser extent, the Nuclear Regulatory Commission. At the state level, these groups were successful in placing several anti-nuclear referendums on the ballot. And, at both levels, environmental groups sought to influence elected representatives by publicizing their voting records on nuclear and environmental matters. In Sweden, anti-nuclear protesters organized themselves as public interest groups. Unlike their American counterparts, they preferred to work directly through the established party system rather than through either the legislature or the bureaucracy. Both the Centre party, whose electoral base


26 For the Swedish nuclear controversy, see Daleus, 'A Moratorium in Name Only'; Nelkin and Pollak, 'The Politics of Participation'; Dean Abrahamsen, 'Governments Fall as Consensus Gives Way to Debate', *Bulletin of the Atomic Scientist*, xxxv (1979), 30–7; Ann-Marie Westmann,
<table>
<thead>
<tr>
<th></th>
<th>Assimilative strategies aimed at political inputs</th>
<th>Assimilative strategies aimed at political outputs</th>
<th>Confrontational strategies against political process</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lobbying/petitioning</td>
<td>Elections/referendums</td>
<td>Interventions in licensing procedures</td>
</tr>
<tr>
<td>United States</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Sweden</td>
<td>high</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>West Germany</td>
<td>low</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>(later: high)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>(later: high)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
is essentially middle-class and whose constituency includes many young, educated people who are attracted to the issues of social decentralization and self-management, and the Eurocommunist Swedish Communist party adopted anti-nuclear positions in the mid-1970s. Other significant arenas of political participation included the several investigative government commissions on nuclear policy that sat during the course of the controversy and the national referendum on the nuclear program in 1980.

By contrast, in political systems where the established channels of political articulation offer few opportunities to voice protest, movements opted for more confrontational ‘outsider’ strategies. For instance, no major party in either France or West Germany adopted a clear-cut anti-nuclear position during the controversy. In both countries parliament exercised next to no control over nuclear policy, and executive agencies were inaccessible to the nuclear opposition, rendering lobbying strategies futile. Finally, neither in West Germany, where the Basic Law rules out plebiscites, nor in France, where political elites stood in the way, could anti-nuclear groups dramatize their demands through national referendums. In France, the Socialists raised the idea of a referendum while in opposition but quickly abandoned it once in office. Given the inaccessibility of the existing political input structures, anti-nuclear movements in both countries began to press for structural change through the new anti-nuclear ecology parties of the late 1970s.

Variations in movement strategies also hold good with respect to the implementation of nuclear policy. In the United States and West Germany, each with weak implementation capacities, an arm’s length relationship between government and the nuclear industry prevails. At least on paper and in the formal decision-making procedures, state regulators are neutral referees in conflicts between the industry and its critics. Licensing procedures, therefore, allowed nuclear critics to intervene in public investigative hearings and to sue against regulatory decisions. However, procedures are cumbersome and involve a plethora of competing agencies. In the United States and to a lesser extent in West Germany, intervention in licensing activities emerged as a major strategy of anti-nuclear groups and legal councillors. Moreover, in


This content downloaded from 212.175.32.136 on Tue, 08 Sep 2015 13:33:52 UTC
All use subject to JSTOR Terms and Conditions
both countries, the courts became a central battlefield in the controversy.\textsuperscript{29} Although court action was frequently responsible for significant delays in nuclear construction projects, the courts were rarely inclined to rule in the plaintiff’s favour.\textsuperscript{30} Thus, anti-nuclear activists realized that licensing skirmishes and litigation could only temporarily stave off nuclear projects and that other means of achieving their cancellation or shut-down were required.

In France and Sweden, in contrast, relatively effective public implementation capacities prevented intervention in the regulatory machinery. Both countries disallow broad political participation in licensing procedures and conduct no public hearings about licensing applications.\textsuperscript{31} The protesters therefore had to look for other promising strategies to alter the course of nuclear policy.

The frequency of confrontational protest behaviour also supports the contention that a link exists between movement strategies and opportunity structures.\textsuperscript{32} Confrontational incidents are most common in the regimes I have categorized as closed, as a measure of demonstration activity in the four countries indicates. In France, between 1975 and 1977, approximately 175,000 people rallied against nuclear power in ten demonstrations. Determined police action against the demonstrators subsequently led to a decline in such mass events. In West Germany, the intransigence of political elites provoked demonstrations too, but a weak state did not act decisively to quell the unrest. From February 1975 to April 1979, approximately 280,000 people participated in seven demonstrations at nuclear sites. Several site occupations were also attempted. In the aftermath of the Three Mile Island accident, in the fall of 1979, approximately 120,000 attended a Bonn demonstration against nuclear power. Several large demonstrations have taken place since then.

In the less intransigent Swedish and American systems, demonstrations have played only a minor role. Two, attended by between 10,000 and 15,000

\textsuperscript{29} The role of litigation in West German and American anti-nuclear activities is discussed in Constance Ewing Cook, Nuclear Power and Legal Advocacy: The Environmentalists and the Courts (Lexington, MA: Lexington Books, 1980); Kitschelt, Kernenergiepolitik, Chap. 5.4; Nelkin and Pollack, The Atom Besieged, Chap. 11.


\textsuperscript{32} These correlations between political structures and protest activity also appear in earlier movements from which anti-nuclear groups recruited some of their participants. Student movements in the late 1960s, for instance, were more militant and embittered in West Germany and France than in Sweden or in the United States.
people from Sweden and Denmark, were held in 1976 and 1977 to protest against the construction of the Barseback nuclear complex. In the United States, demonstrations and civil disobedience were strategies ‘imported’ from Western Europe. Despite its head start in the late 1960s, the American anti-nuclear movement staged its first large-scale demonstration only in 1978, at the site of the Seabrook plant in New England.\(^{33}\) To be sure, the accident at Three Mile Island in March 1979 was followed by a number of demonstrations with large turnouts, such as those in New York City, Washington, DC, and San Francisco, but occurring as they did, in the wake of what was depicted as a near national disaster, they must be viewed as temporary aberrations from the prevailing American pattern of assimilative protest.

Overall, there is convincing evidence that political opportunity structures direct the paths of social mobilization taken by the various national nuclear protest movements. It is a different matter, however, whether even strategies well-adapted to differing political regimes necessarily produce success in terms of procedural, substantive or structural impacts.

**Political Opportunity Structures and Social Movement Impacts**

If political opportunity structures shape the impact of anti-nuclear movements on policy, we should not expect policy impacts to be attributable to the overall scale and intensity of protest but rather to vary, within limits, independently of them. High mobilization does not necessarily lead to profound impacts if the political opportunity structures are not conducive to change. Conversely, lower mobilization may have a disproportionate impact owing to properties of the political opportunity structure.

To test this argument, one would ideally need a good common measure of overall movement strength in each country. However, given the numerous protest strategies adopted, such a measure is difficult to find. Poll information on opinions about nuclear energy provides a very rough indicator; but polls are extremely sensitive to the way questions are phrased and are, at best, indirectly related to anti-nuclear protest activities.\(^{34}\) Moreover, poll findings are not stable over time and are vulnerable to a public issue-attention cycle.\(^{35}\)

That said, and allowing for the lack of strictly comparable cross-national sur-

\(^{33}\) The Seabrook controversy is discussed in Steven Barkan, ‘Strategic, Tactical and Organizational Dilemmas of the Protest Movement Against Nuclear Energy’, *Social Problems*, xxvii (1979), 19–37; and in Harvey Wasserman, *Energy War: Reports from the Front* (New York: Lawrence Hill, 1979).

\(^{34}\) Survey questions have been manipulated so as to create a virtual ‘politics of nuclear polling’, with advocates and opponents of nuclear power using the surveys most favourable to their own position. This is discussed in Otwin Renn, *Kernenergie aus der Sicht der Bevölkerung* (Jülich: Kernforschungsanlage Jülich, 1977), pp. 47–9.

veys, several tentative generalizations can be drawn from opinion surveys taken during the late 1970s, at the height of the public controversy. Anti-nuclear sentiments appear to have peaked once around 1976 and again in 1979, after the Three Mile Island accident, France being an exception in the latter instance. At these peaks, a plurality of respondents in all countries favoured nuclear power (35 per cent to 50 per cent of respondents), sizeable minorities opposed it (30 per cent to 45 per cent), and significant, but over time declining, groups expressed no opinion (10 per cent to 30 per cent). Anti-nuclear preferences reached a plurality only for a brief period following the Three Mile Island accident. Since 1979, they have declined in all four countries. Survey findings as similar as these can clearly shed little light on the significantly different policy impacts of anti-nuclear protests in France, West Germany, the United States and Sweden.

More direct measures of protest mobilization also do not yield plausible associations with movement impacts. The number of participants in the various protest activities, for instance, even when standardized for country size, turns out to be a poor predictor of movement impact. The United States has had a comparatively low level of mobilization, but its nuclear program is stalemated. France, in contrast, has had much greater mobilization, but its program has experienced little disruption. Also, the number and total membership of anti-nuclear protest organizations does not yield a reliable independent measure of protest intensity, because cross-national variations reflect varying opportunity structures rather than varying intensities of mobilization. This caveat is borne out in the open Swedish and American regimes, where protests are more formally organized and rely heavily on established nature-protection lobbies. These lobbies are much less important in France and West Germany.

Any assessment of the overall strength of the anti-nuclear movement must further consider the complication that ‘power’ and ‘strength’ are relational concepts, which measure not only the resources and activities of anti-nuclear groups but also those of the pro-nuclear advocates. If, for instance, governments had changed their evaluation of nuclear power autonomously, not merely as a reaction to the more or less effective veto-power of anti-nuclear protesters, then less pro-nuclear policies would indicate an intrinsic weakening of the pro-nuclear advocates, not the strength of anti-nuclear protest or opportunity structures conducive to the exercise of veto-power. Empirically, though, there is little evidence for autonomous change in the positions, preferences and resources of the nuclear advocates in the four countries during the protest period. Everywhere, nuclear manufacturing industries, electric utilities and state agencies promoting nuclear power remained firmly committed to the

36 For opinion surveys about nuclear energy issues, see Gerard Duménil, ‘Energie nucléaire et opinion publique’, pp. 317–74 in Fagnani and Nicolon, Nucleopolis (France); Barbara Farah et al., Public Opinion About Energy: A Literature Review (Golden, CO: Solar Energy Research Institute, 1979) (United States); Renn, Kernenergie aus der Sicht der Bevölkerung and Wahrnehmung und Akzeptanz technischer Risken (Jülich: Kernforschungsans Jülich, 1981) (Germany); and Hans Zetterberg, The Swedish Public and Nuclear Energy (Sweden).
new technology. For example, industrial policies in both Sweden and France targeted nuclear plant manufacture as a major export industry and, therefore, it enjoyed a secure political position. The wholly or partially nationalized electric utilities were also strong lobbyists. Nevertheless, the outcome of the nuclear controversy in the two countries differs dramatically with respect to the long-term future of their nuclear industries.37

Overall, the evidence about mobilization does not support an explanation of differential movement impact based solely on the relative internal strengths and weaknesses of the movements and their pro-nuclear opponents. The alternative to this approach is to try to capture the dynamic interplay between movement mobilization and regime response by examining systematically the types of impacts anti-nuclear movement have had. To this I now turn.

PROCEDURAL IMPACTS

Anti-nuclear movements have made procedural inroads when they have been able to gain greater access to formal political decision-making. At one end of the spectrum, procedural impacts in France have been virtually non-existent. Because the French party system is organized along a bipolar socio-economic cleavage, its parties have been reluctant to represent anti-nuclear demands. This is well illustrated by the inability of either the Socialists or the Communists to respond to the protest. The Socialists, for example, temporarily flirted with the anti-nuclear cause in the 1978 and 1981 campaigns but once in government quickly backed away. Vacillation also characterized the position of the Socialist-leaning labour union, whose leadership was sceptical about the merits of nuclear power but was unable to generate widespread support for an anti-nuclear position. The Communist party and its labour union, which is firmly entrenched in the utility and energy industries, were even less accessible.

The anti-nuclear movement also met with indifference and worse from the various state authorities. During the Giscard administration, for instance, the nuclear issue was never discussed at length in the French legislature. Once elected, the Socialist government rid itself of the internally divisive issue in an early and brief parliamentary debate in October 1981. At that time, the new government simply decided to continue the nuclear policy of its conserva-

37 It has also been argued that weakness of a nation’s energy sector, above all the absence of strong oil companies, explains why governments protect nuclear and other energy policies more from movement challenges than do countries with strong, indigenous energy industries. For this argument, see Peter Gourevitch, ‘The Second Image Reversed: The International Sources of Domestic Politics’, International Organization, xxxii (1978), 881–911, esp. p. 906. At first blush, this model seems to explain the differences between France and the United States. The weak French energy sector requires firm government support, whereas the United States can afford a more pluralist style because its energy sector is strong and can fight for itself. The model fails to explain, however, why West Germany and Sweden, each with comparatively weak energy industries, were unable to imitate the French strategy and, instead, retreated, each in its own particular way, from an all-out, long-term commitment to nuclear power.
tive predecessor and gave little consideration to the dissenting minority within its own parliamentary party. Furthermore, no efforts were made to represent anti-nuclear interests in arenas of quasi-corporatist decision-making or to organize a national plebiscite to resolve the conflict.\textsuperscript{38} The French anti-nuclear movement also failed to gain access to executive branch agencies and, although the government streamlined nuclear licensing procedures and made them more comprehensive in the late 1970s, the new rules still prevented the opposition from participating in public licensing hearings and from appealing effectively to the courts.\textsuperscript{39} Finally, the French state responded to anti-nuclear demonstrations and civil disobedience with a dose of retaliation so heavy – as in its response to the 1977 demonstration against the new fast breeder reactor in Malville – that anti-nuclear activities have since been effectively discouraged.

The anti-nuclear movement in West Germany did not fare much better with the established political parties than did its French counterpart. The conservative opposition parties were clearly in favour of nuclear power. The German labour unions were also strongly supportive of the expansive nuclear program. And although the parties in the Social Democratic–Free Democratic coalition government were internally divided between pro-union and pro-business nuclear advocates and important oppositional minorities, a situation that led to a temporary policy stalemate, they never unequivocally represented the anti-nuclear position. Moreover, the parties were unable to influence the firmly pro-nuclear policies of the Social Democratic–Free Democratic government. It is therefore not surprising that parliamentary debates on nuclear energy during the period never effectively gave voice to the anti-nuclear position. A parliamentary commission on nuclear energy which included pro- and anti-nuclear ‘experts’ was convened in 1979, late in the controversy. It issued a strategically ambivalent mid-term report in 1980, which pro-government partisans hoped would draw anti-nuclear sympathizers over to the government parties in that year’s national election.\textsuperscript{40} After the election, in 1982, however, a broad majority of commissioners endorsed the planned nuclear program, almost in its entirety. Overall, the anti-nuclear movement in West Germany made no gains in procedural representation, for its mobilization

\textsuperscript{38} The new Socialist government allowed consultative local referendums on nuclear power projects. But this provision was far less sweeping than it sounds, for referendums may be overruled by decisions of regional political bodies, and the reform was accompanied by government and electric utility threats of economic hardship for uncooperative regions. The licensing reform by the Socialist government is described in M. Rappin, ‘Dezentralisierung des französischen Genehmigungsverfahrens’, Atomwirtschaft-Atomtechnik, xxvii (1982), 39–41.

\textsuperscript{39} For the litigation initiated by French anti-nuclear activists, see Colson, \textit{Le Nucléaire sans les Français}, pp. 139–50. and Nelkin and Pollak, \textit{The Atom Besieged}, Chap. 11.

\textsuperscript{40} The political dynamics of this commission are analysed in Herbert Kitschelt, ‘Der Zwischenbericht der Enquete-Kommission “Zukünftige Kernenergiepolitik”: Stagnation oder Innovation in der politischen Ökonomie des westdeutschen Energiesektors?’ \textit{Jahrbuch Technik und Gesellschaft}, 1 (1982), 165–91.
failed to open any new party, legislative, corporist or (constitutionally forbidden) plebiscitari

The obstacles encountered by the West German nuclear program were in fact generated at the implemen
tation end of the policy process, by procedures that were neither fully open nor closed to public participation. The exist
ing licensing procedures were fragmented and did slow the program down, but opponents were not able to use these weaknesses purposively to pursue their own policy agenda because they were unable to extend their procedural participation, e.g., through more extensive citizens’ rights to sue collectively against industrial projects (Verbandsklage). Until 1981, however, when the movement began to wane, the government was unable to take the decisive measures necessary to tip the scales firmly in favour of the program’s proponents by, for instance, streamlining the licensing procedure and restricting the opportunities for procedural obstruction that the anti-nuclear activists enjoyed. By neither consistently repressing anti-nuclear protesters nor granting them new democratic rights, the state may have unwittingly fuelled the movement’s mobilization and thwarted the nuclear program.

The United States, while also a case of policy stalemate, exhibits opportunity structure features that distinguish it from West Germany. America’s ‘decom
pounded’ party system began to assimilate anti-nuclear demands with relative ease in the early 1970s. Then, in 1976, the anti-nuclear movement succeeded in placing anti-nuclear referendums on the ballot in a number of states. This action and the electoral response were measures of public opinion that further sensitized legislators to the issue. Partly as a consequence, more members of Congress, regardless of party affiliation, shifted to an anti-nuclear position. Anti-nuclear activists also strengthened their position in the executive branch. During the Carter presidency, the zenith of the conflict, several anti-nuclear activists were appointed to high-ranking positions in energy and environmental agencies. Even so, they could not conquer entrenched pro-nuclear bastions. The intensifying nuclear power debate led instead to a progressive fragmentation of political power and a stalemate in both Congress and the administra
tion, which prevented adoption of any coherent nuclear policy. Numerous changes in the organization and jurisdiction of legislative and administrative bodies in the nuclear arena illustrate this process.41 Few new policy initiatives were undertaken, and key political actors were unable to forge lasting and effective coalitions. With respect to the reform of nuclear licensing procedures, for instance, neither advocates nor opponents of nuclear power were able to make decisive gains. Although pro-nuclear forces in the United States have regained strength during the Reagan presidency, the stalemate has not

41 There are several overviews of American nuclear energy policy that place it in the more comprehensive setting of American energy policy: Irwin C. Bupp and Jean-Claude Derian, Light Water: How the Nuclear Dream Dissolved (New York: Basic Books, 1978), Chaps. 8 and 10; Walter Rosenbaum, Energy, Politics and Public Policy (Washington, DC: Congressional Qua
been broken and no stable governance of nuclear energy is in sight. The American political opportunity structure has thus facilitated the partial, though inconsistent, inclusion of the anti-nuclear opposition in decision-making arenas.

At the end of the input spectrum furthest removed from France lies Sweden, which exhibits the greatest degree of procedural responsiveness. The anti-nuclear opposition there was represented by two opposition parties (the Communist and the Centre) as early as 1973, and parliament served as a forum for the nuclear debate. Moreover, the pro-nuclear Social Democratic government financed a broadly participatory national debate on the issue. This debate especially had dramatic consequences for the Swedish nuclear energy program, for it intensified and crystallized the public’s anti-nuclear sentiments, which in turn contributed to the defeat of the Social Democrats in the 1976 election. The new government, headed by a prime minister opposed to nuclear energy, was, however, divided on the issue. The government tried at first to end the deadlock by adopting the quasi-corporatist strategy of appointing an energy commission, which was staffed by the chief antagonists in the controversy. But when this effort failed, along with efforts to negotiate a viable policy compromise between the ruling parties, the government collapsed in 1978. The 1979 accident at Three Mile Island injected a new urgency into the debate, which prompted all parties to agree to a national referendum, to be held after the upcoming election, that would remove the issue from the realm of ordinary political campaigning. In summary, the Swedish anti-nuclear movement found that its political opportunity structure offered an eclectic variety of participatory avenues: electoral, corporatist and plebiscitarian. What it did not find, however, were similar opportunities to influence implementation – for example, to intervene in nuclear plant licensure proceedings. In this respect, Sweden is much like France.

SUBSTANTIVE POLICY IMPACTS

What kind of policy impacts have anti-nuclear movements sought? On the one hand, anti-nuclear activists have sought the suspension of nuclear power plant licensing and construction, and, in certain instances, the shut-down of already-operating plants. On the other hand, they have called for a reorientation of energy policies towards energy conservation and research on renewable energy resources. While these goals have been pursued by activists in all of the four countries under consideration, the degree to which they have been successfully pursued varies widely. It is to this variation that we now turn.

That at least several of the movements have been rewarded by scaled-down nuclear programs is clear from Tables 4 and 5, which provide, respectively, information about the number of commercial nuclear power plants under

construction or already in operation in each country during the period from 1974 to 1984, and information about the number of planned nuclear stations not yet under construction. Not surprisingly, the regimes most tolerant of the anti-nuclear opposition, Sweden and the United States, are also the ones with a steady or declining number of plants planned and built in the last decade. West Germany's program is also, if only temporarily, restrained. And, of the four, only France's program continues to grow and to grow rapidly. But what specific attributes of political opportunity structures have made some movements more successful than others in achieving this portion of the anti-nuclear agenda?

One might be tempted to suppose that a simple drop-off in demand for additional electricity explains the differences between the four countries. But growth of electricity demand has also slowed down dramatically in France, where the nuclear program has continued apace. German and American utilities still deplore the expected 'shortfall' of, in their view, potentially inexpensive nuclear base load capacity. Moreover, in all four countries, utilities did not expect declining growth rates in the 1970s and planned many new nuclear

---

### TABLE 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>10</td>
<td>30</td>
<td>44</td>
<td>54</td>
</tr>
<tr>
<td>Sweden</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>United States</td>
<td>87</td>
<td>136</td>
<td>140</td>
<td>125</td>
</tr>
<tr>
<td>West Germany</td>
<td>15</td>
<td>18</td>
<td>19</td>
<td>21</td>
</tr>
</tbody>
</table>

* The higher figure includes plants under construction, but temporarily mothballed (1984: 9) or permanently abandoned (1984: 10).


---

### TABLE 5

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>15</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>United States</td>
<td>125</td>
<td>64</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>West Germany</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

* Figures in brackets include plants without definite construction schedule.

Sources: as for Table 4.
stations. In France, however, a political–economic regime intransigent to anti-nuclear activists was able to realize such plans and overbuild nuclear capacity to an extent that it precipitated a financial crisis of the nationalized Electricité de France.\textsuperscript{43} In West Germany, the United States and Sweden, effective anti-nuclear opposition ‘saved’ utilities from making investments that would have appeared uneconomic in retrospect. Why did this happen?

Where political opportunity structures were conducive to popular participation, anti-nuclear activists could impose economic penalties on nuclear builders, by slowing the construction of plants being built and increasing the risk of future investments. In the United States, Sweden and West Germany, anti-nuclear activists were continually able to raise the costs of plants, but those in France were unable to do this. Greater responsiveness to the anti-nuclear opposition invariably led to extremely tight and often changing safety regulations.\textsuperscript{44} Once formulated, these new safety standards allowed opponents to intervene to insist that they be complied with. The two factors reinforced each other; when nuclear regulatory agencies tightened their safety standards, opponents felt justified in their suspicions and pressed for additional requirements or else requested that existing plants be upgraded to meet the latest standards. Thus, new safety standards and the delays they brought with them – both resulting from relatively open political opportunity structures – increased the capital costs and finance charges on borrowed capital incurred by the builders of nuclear plants.

Table 6 shows that construction schedule delays were most pronounced in the United States and West Germany, both of which have fragmented implementation structures.\textsuperscript{45} Here, in addition to licensing procedures, the courts also contributed to the delay problem when they suspended construction work during litigation. Much shorter delays were typical in France and Sweden, where tight implementation procedures offer few opportunities for outside intervention. The contrast is best illustrated with a comparison of average completion times; in the United States, it takes twelve to fourteen years to complete a commercial nuclear plant, while in France it takes only six.

In Sweden, nuclear policy was changed not by disrupting the policy


\textsuperscript{44} This process was accompanied by increasing outlays for nuclear safety research that led to further regulatory requirements and delays of nuclear power plants. See Barry Weingast, ‘Congress, Regulation and the Decline of Nuclear Power’. \textit{Public Policy}, xxviii (1986), 231–55.

\textsuperscript{45} These data would be even more striking if only plants originally scheduled for completion between 1970 and 1980 had been included. By the 1980s, after the controversy’s peak, plants were delayed an average of 73.7 months in the United States, 42.2 months in the Federal Republic, 15.9 months in France, and 17.2 months in Sweden. By 1984, some time after the peak of the nuclear controversy, delays for this group of power plants had increased still further: 86 months in the United States, 56 in West Germany, 26 in Sweden, and 16 in France. Data are calculated according to sources and procedures described under Table 6.
TABLE 6  Average Construction Delays of All Nuclear Power Plants Under Construction or in Commercial Service (Months)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>0.7</td>
<td>3.6</td>
<td>7.1</td>
<td>11.3</td>
</tr>
<tr>
<td>Sweden</td>
<td>2.7</td>
<td>4.9</td>
<td>15.9</td>
<td>19.8</td>
</tr>
<tr>
<td>United States</td>
<td>20.0</td>
<td>35.9</td>
<td>49.4</td>
<td>53.1</td>
</tr>
<tr>
<td>West Germany</td>
<td>6.1</td>
<td>13.8</td>
<td>30.6</td>
<td>42.4</td>
</tr>
</tbody>
</table>

* Delays for each plant were measured as number of months behind construction schedule expected at that time when the plant order was given. For each country, delays were calculated only for plants already in operation or under construction and still scheduled to be completed. Plants where construction has not yet begun or plants mothballed while under construction are not included. Sources: calculations based on sources given in Table 4.

implementation process but by working through the ‘input side’ of politics. As a result of a difficult and long drawn-out process of mutual adjustment among opposing groups, not a single political party continued to advocate further expansion of the nuclear program when the national referendum was held in March 1980. Although anti-nuclear activists did not succeed in persuading a majority of the voters to support an immediate halt to all construction activity and a dismantling of existing plants, the Swedish government has taken the magnitude of public opposition into account and has ordered no new plants, a move that is certain to guarantee the demise of the Swedish nuclear industry. The strength of the ‘input side’ of Swedish politics is also demonstrated by that country’s ability to respond positively to the other half of the anti-nuclear agenda, which calls for an energy policy orientated towards conservation and the development of renewable fuels, especially biomass.

Why such a basic shift in overall energy strategy has not occurred in either the United States, West Germany or France is again to be traced back to variations in national opportunity structures. Energy conservation programs, for instance, working through incentives, taxes, regulation and state investment in infrastructure have been most aggressively pursued in the ‘high intervention’ political economies of Sweden and France. In Sweden, the trade-off between nuclear power and conservation has been direct. In France, however, the commitment to conservation complements the government’s existing one to nuclear power and, at least indirectly, is attributable to elite efforts to appease the anti-nuclear opposition without yielding to its key demands. In the United States and West Germany, conservation policies have scarcely been pursued beyond allowing free market mechanisms to increase the prices of scarce resources.


Government research in energy conservation and renewable energy technologies exhibits similar patterns. While governments in all four countries targeted their financial support almost exclusively on nuclear research until the mid-1970s, funding of new energy technologies managed to take off in the two open regimes. In contrast to Sweden, the research program in the United States has been beset with problems, most of which stem from this new research area’s unstable supporting coalition and to inefficient program implementation.\(^{48}\) Using budget outlays for nuclear and renewable energy technologies as an indicator, Table 7 illustrates the differences between the energy technology policies of the four countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Nuclear fission</th>
<th>Renewables/energy conservation</th>
<th>Overall energy technology expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>0.14</td>
<td>0.65</td>
<td>1.05</td>
</tr>
<tr>
<td>United States</td>
<td>0.49</td>
<td>0.36</td>
<td>1.61</td>
</tr>
<tr>
<td>France</td>
<td>0.72</td>
<td>0.28</td>
<td>1.54</td>
</tr>
<tr>
<td>West Germany</td>
<td>0.83</td>
<td>0.12</td>
<td>1.39</td>
</tr>
</tbody>
</table>


Is it not possible to explain all the changes in nuclear and energy technology policy through shifting electoral fortunes and changes of governments, rather than through the more stable political opportunity structures? For France and Germany, the answer must be no. Changes from a conservative–liberal to a socialist government and vice versa made little difference to energy policy. Even in the United States, four different presidents (Nixon, Ford, Carter and Reagan) declared themselves more or less enthusiastically in favour of nuclear energy, but none was able to create an effective coalition to support it. Generally, long-term energy strategies appear to be difficult to maintain in the fluid American system of policy formation, as both the rapid rise and demise of renewable energy research and of the nuclear fast breeder reactor technology demonstrate.

The only case broadly consistent with the importance of elections and changes of government is that of Sweden, where a government change in

\(^{48}\) An instructive analysis of the budget decisions for solar energy by the US Congress is provided by W. Henry Lambright and Albert Teich, ‘Policy Innovation in Federal Research and Development: The Case of Energy Research and Development’, *Public Administration Review*, 41.1 (1979), 140–7. A detailed comparative analysis of the formation and implementation of energy technology policies can be found in Kitschelt, *Politik und Energie*, Chap. 6.
1976 did precipitate a shift in energy policy. However, this case is not inconsistent with our competing regime hypothesis, for the latter would also predict policy changes to occur as a consequence of electoral and government changes whenever political regimes are open and have the capacity for effective policy implementation. But, even in Sweden, electoral politics is of only limited significance for policy innovation because policy changes are frequently built on much broader than minimum winning coalitions, as evidenced by the unanimous decision of the political parties in 1980 not to pursue nuclear power beyond the twelve-reactor program.

Differences in nuclear policy among the four countries are also not entirely explained by the import dependence argument, which predicts intransigent pursuit of nuclear energy whenever dependence is high. Thus, the French, Swedish and West German dependence levels on imported oil are too similar to justify the significant differences that can be found in their respective nuclear policies. Moreover, in Sweden uranium deposits are the only significant indigenous non-renewable fuel reserve. Despite this, Sweden is attempting to withdraw from the nuclear economy. In the American case, energy import sensitivity and vulnerability is much lower than in Europe, but the United States’ role as the hegemonic Western power and the absolute magnitude of American energy imports in world trade render the long-term question of energy supply no less significant in the United States than in Europe.

Raw figures about foreign energy dependence and national energy resources are, by themselves, not likely to explain public policy. A shortcoming of both resource dependence and change of government explanations is that they are not sophisticated enough to reconstruct the actual process of nuclear power policy formation. Opportunity structures come much closer to explaining the process through which a new policy is learned or an old policy is reaffirmed in the face of challenging political demands.

**Structural Impacts of Anti-Nuclear Movements**

Aside from the procedural and substantive impacts, the impact of anti-nuclear movements on political regimes themselves may vary between countries, depending on the type of opportunity structure that exists. The less innovative and more immobile a political regime, the greater the risk that this inflexibility itself will trigger demands that go beyond the immediate policy issue to ones threatening the legitimacy of the regime.

Changes in political implementation capacities resulting from anti-nuclear protest are difficult to detect in the four countries for the 1973–83 period.

---

49 Large German coal deposits do not improve this picture dramatically. Because mining and burning coal have deleterious environmental and economic consequences, Germany has been hesitant to exploit this resource at an accelerated pace.

Everywhere, the inertia of administrative institutions and the economic power of the established actors in the energy sector are formidable. Structural impacts do, however, stand out with respect to the four countries’ input patterns. Where the political input structures are closed, noteworthy efforts have been made to realign the party system. New ‘green’ or ecological parties have appeared and adopted the nuclear issue as a major plank in their political programs. These parties tend not to be organized along traditional cleavages of class, religion and ethnicity. Instead, they are mobilized on the basis of alleged inequalities of ‘qualitative’ and ‘reproductive’ life chances, which are created by the subordination of nature and society to large-scale economic enterprise and bureaucratic state institutions.

We would expect ecological parties to be stronger in regimes that are less responsive to anti-nuclear demands, such as those of France and West Germany, and weaker or non-existent in more innovative ones, such as those of Sweden and the United States. In fact, the West German ecology party, *Die Grünen*, founded in the late 1970s, has managed to win more than 5 per cent of the vote in most West German state elections since 1979. In the March 1983 federal election, for example, it received 5.6 per cent of the popular vote, and in the European election of 1982 8.2 per cent. In France, ecological voting lists experienced remarkable successes in the local elections of 1977. They were less successful in the 1978 National Assembly elections, however, when they could not agree on either participation in the election or a common campaign strategy. Nevertheless, the ecological candidate in the French presidential election of 1981, Brice Lalonde, received almost 4 per cent of the vote on the first ballot. The two ecological parties participating in the 1984 election together won 6.7 per cent of the vote.

In contrast, green parties in the United States and Sweden have received little support. The Swedish ecological party failed to receive 2 per cent of the vote in the Riksdag election of 1982 and has remained well below the minimum 4 per cent threshold it needs for representation in parliament. The closest equivalent to an ecology party in the United States, the Citizens’ party, is insignificant.

Although there is a striking correlation between the performance of ecological parties and the outcomes of the nuclear conflict in the early 1980s, one

---


52 The new parties thus destabilize the formation of cleavages that have been institutionalized in West European party systems throughout most of this century. Compare Seymour Martin Lipset and Stein Rokkan, eds, *Party Systems and Voter Alignments* (New York: Free Press, 1967).

53 This argument is elaborated in the theories referred to in fn. 3.
is well advised not to interpret the new parties as ‘single issue’ formations. Thus, the nuclear conflict and political opportunity structures may offer a good explanation of the emergence of such parties in a specific historical period, but a multitude of other factors may influence their future course.54

CONCLUSION

This comparison of nuclear power conflicts in four countries shows that the mobilization strategies and impacts of social movements can, to a significant degree, be explained by the general characteristics of domestic political opportunity structures. Furthermore, the cases show that governments do not necessarily engage in a reactive process of learning when faced with unexpected opposition to a policy. In the case of nuclear energy, the capacity to learn from the experience of manifest conflict did not simply follow from the magnitude of protest. Rather, it was shaped in certain pre-established ways by the channels and opportunities that political regimes offered to opponents to disseminate their message and disrupt established policies. Variations of such institutional rules led to different dispositions of governments to defend or revise policies. Where political input structures were open and responsive to the mobilization of protest, as in Sweden and to a lesser extent in the United States, a search for new policies was triggered. Where they were closed, as in France and West Germany, governments insisted more intransigently on a predetermined policy course. Where state capacities to implement policies were weak, as in the United States and West Germany, the nuclear protest movement had at least a chance to disrupt the policy against which it was mobilized. Where political capacities were stronger, as in Sweden and France, nuclear policy was shielded from most of the attacks on its implementation. The combination of political input and output structures in each country sets limits on policy innovation. Where openness was high and capacity strong, innovation tended to be greater. Sweden approximates this configuration best. Where the reverse configuration existed, policy-making immobility prevailed. This is especially patent in the West German case, and to a lesser extent in the United States. Under these conditions, established policies were stalemated, and new policies could not be agreed upon or implemented.

Theories are fruitful only if they can be applied to cases beyond the ones they were first designed to explain. In extending the logic of the present argument to nuclear power conflicts in other countries, one must, however, take into account the possibility that several factors, controlled in this study, might vary in a larger sample of cases, most notably the relative size of nuclear

54 Thus, ecological parties in countries with weak nuclear conflicts or with opportunity structures not conducive to the formation of new parties will benefit in the future from a demonstration effect provided by the successful ecological parties, especially the West German party. A more exhaustive comparative analysis of ecological parties in different countries would require a detailed examination of the socioeconomic development, the political culture and the system of party competition in each instance.
programs and the intensity of anti-nuclear mobilization. This qualification underlines again the caveat that the regime hypothesis does not explain differences of social movement mobilization and energy strategies in their entirety.

Beyond the nuclear case, it is conceivable that the regime hypothesis can explain the strategies and impacts of other movements concerned with qualitative life-chances and the physical structuring of the social environment. Many of these conflicts cut across the social cleavages currently institutionalized in party systems, arenas of functional interest group representation and administrative agencies. Political regimes have a varying propensity to innovate in procedural, substantive and structural ways when confronted by challenging new groups. Energy policy aside, environmental protection, consumer safety regulation, information systems control, genetic engineering regulation, military technology and strategic planning, urban development and transportation planning are but a few examples of issue areas that may precipitate new social movements with dynamics similar to those of anti-nuclear movements.