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.

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: (a) identifying novel political and social phenomena and systematically describing them; (b) evaluating prior explanatory hypotheses, discovering new hypotheses, and assessing these new causal claims; (c) gaining insight into causal mechanisms; and (d) 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.5

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?6

Identifying evidence that can be interpreted as diagnostic depends centrally on prior knowledge.7 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 merits 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.

**Recurrent Empirical Regularities.** These are established patterns8 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 not 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.10

Mahoney (2010, 127–28) illustrates descriptive inference in process tracing with Tannenwald’s (1999) study of the “Nuclear Taboo.”11 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 policy makers (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.12 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 pervasiveness 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, 31–32), 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></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NECESSARY FOR</strong></td>
<td><strong>YES</strong></td>
<td><strong>NO</strong></td>
<td><strong>YES</strong></td>
</tr>
<tr>
<td><strong>AFFIRMING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CAUSAL INFERENCE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUFFICIENT FOR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AFFIRMING CAUSAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>INFERENCE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Passing:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>affinity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>b. Failing:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Yes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Passing:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>affinity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Study</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>c. Implications</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>for rival hypotheses:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passing:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly weakens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>them.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failing:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly strengthens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>them.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>d. Failing:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eliminates hypothesis.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Teacher: Understanding Process Tracing

Table 2

Overview of “Silver Blaze”

<table>
<thead>
<tr>
<th>Causal Puzzle</th>
<th>Main Characters</th>
<th>Hypotheses</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>Silver Blaze, the racehorse that is the favorite for the Essex Cup, has disappeared.</td>
<td>Silver Blaze.</td>
</tr>
<tr>
<td>John Straker, the horse’s trainer, has been killed by a terrible blow that shattered his head.</td>
<td>Fitzroy Simpson, a prime suspect, has been lurking around the stable seeking inside information about the race.</td>
<td>H1. Romantic entanglement</td>
</tr>
<tr>
<td>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.</td>
<td>Colonel Ross is the owner of King’s Pyland Stables and of Silver Blaze.</td>
<td>H2. Chain of events started in Straker household</td>
</tr>
</tbody>
</table>

Independent Variables | Intervening Variables | Dependent Variables |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>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>

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 points to the possibility that someone known to the dog—i.e., Straker—abducted the horse (H3), 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.
 unlikely would assume, as just stated, that a timid individual such as Simpson Hence, he fails this second hoop test. Table 4 also illustrates alteration non-menacing appearance might seem to preclude his being a possibility. This provides a ever, those with no smoking gun may not be innocent. In other words, this suggests, but hardly confirming, intent to harm. It might possibly be a coincidence. alternative interpretation that the knife was somewhat unusual, it is a straw-in-the-wind that makes \( H_4 \) more plausible, without confirming it.

### Table 4  
**Hoop Tests**

<table>
<thead>
<tr>
<th>( H_6 )</th>
<th>Simpson killed Straker.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inference.</strong></td>
<td>This weapon is consistent with the hypothesis, but does not by itself demonstrate Simpson’s guilt.</td>
</tr>
<tr>
<td><strong>Summary.</strong></td>
<td>Simpson had a potential weapon, so ( H_6 ) passes this hoop test.</td>
</tr>
<tr>
<td>( H_6 )</td>
<td>Simpson killed Straker.</td>
</tr>
<tr>
<td><strong>Clues.</strong></td>
<td>Simpson’s timid, non-menacing appearance, plus the fact that Straker’s “head had been shattered by a savage blow from some heavy weapon.”</td>
</tr>
<tr>
<td><strong>Inference.</strong></td>
<td>With a stronger assumption based on his appearance, Simpson could not have inflicted the blow that shattered Straker’s head.</td>
</tr>
<tr>
<td><strong>Alternative Inference.</strong></td>
<td>With a weaker assumption, Simpson’s appearance raises doubts that he would have committed the murder, but does not preclude it.</td>
</tr>
<tr>
<td><strong>Summary.</strong></td>
<td>With a stronger assumption this is a hoop test which ( H_6 ) fails; with a weaker assumption it is a straw-in-the-wind test which casts doubt on ( H_6 ).</td>
</tr>
</tbody>
</table>

Table 4 presents two hoop tests, both focused on the hypothesis that Simpson killed Straker (\( H_6 \)). 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 nonmenacing person would commit such a murder. This would yield a straw-in-the-wind that casts doubt on the idea that he is the murderer—yet he remains a possible suspect.

### Table 5  
**Smoking-Gun Tests**

<table>
<thead>
<tr>
<th>( H_2 )</th>
<th>The chain of events started in Straker’s household.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clues.</strong></td>
<td>The household maid brought the stable boy curried mutton, and he was found later to have been drugged.</td>
</tr>
<tr>
<td><strong>Inference.</strong></td>
<td>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>
</tr>
<tr>
<td><strong>Summary.</strong></td>
<td>The clues yield a smoking-gun test that confirms ( H_2 ).</td>
</tr>
<tr>
<td>( H_4 )</td>
<td>Straker planned to harm the horse.</td>
</tr>
<tr>
<td><strong>Clue.</strong></td>
<td>Unusual, surgical knife found with Straker.</td>
</tr>
<tr>
<td><strong>Inference.</strong></td>
<td>The knife is interpreted as exceptionally unusual—establishing intent to harm.</td>
</tr>
<tr>
<td><strong>Alternative Inference.</strong></td>
<td>The knife is interpreted as somewhat unusual, suggesting, but hardly confirming, intent to harm. It might possibly be a coincidence.</td>
</tr>
<tr>
<td><strong>Summary.</strong></td>
<td>If the knife is exceptionally unusual, it is a smoking gun that confirms ( H_4 ). With a weaker interpretation that the knife was somewhat unusual, it is a straw-in-the-wind that makes ( H_4 ) more plausible, without confirming it.</td>
</tr>
</tbody>
</table>

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 Simpson’s guilt, 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 (\( H_2 \)). By contrast, the hypothesis that Straker planned to cause harm (\( H_4 \)) 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 (\( H_6 \)) and Straker (\( H_7 \)) 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 (\( H_8 \)), 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, \( H_8 \), 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
come tests 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 (2010a) 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 (2005).

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 (2004a, 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, chaps. 2–3). 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, chap. 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 (2004a, 198–99; 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 discussion seems descending. Further, the female interviewee 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-à-vis quantitative in terms of four dimensions, see Collier, Brady, and Seawright (2004a, 177–82).

14. Another puzzle is explaining the disappearance of the horse, but as Holmes himself emphasizes, that is a secondary issue (see p. 31 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 causality, are true. See, for example, Freedman (2006b).

17. In one story, Sherlock Holmes takes a strong stand on coincidences (“Adven- ture 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 coinci- dence. 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, 331–32).

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


REFERENCES


