Henry Brady and David Collier, eds, *Rethinking Social Inquiry* Rowman and Littlefield, 2010 (second edition)

Chapter 10

Process Tracing and Causal Inference

Andrew Bennett

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.1 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

Maria Gould, Jody La Porte, and Miranda Yaver provided valuable comments on an earlier draft of this chapter.

1. Good examples include Drezner (1999), Eden (2004), George and Smoke (1974), Homer-Dixon (1999), Khong (1992), Knopf (1998), Larson (1997), Moravcsik
(1998), Owen (1997), Rock (1989, 2000), Sagan (1993), Shafer (1988), Snyder (1984, 1991), Walt (1996), and Weber (1991). Brief descriptions of the research designs employed by Drezner, George and Smoke, Homer-Dixon, Khong, Knopf,

Larson, Owen, Sagan, Shafer, Snyder, and Weber are provided by George and Bennett
(2005: 118–119, 194–97, 302–325).

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 case2 that contribute to supporting or overturning alternative explanatory hypotheses. A central concern is with sequences and mechanisms in the unfolding of hypothesized causal processes.3 The researcher looks for the observable implications of hypothesized explanations, often examining 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 above 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 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 (a)

	Sufficient To Establish Causation (b)	
	No	**Yes**
Necessary to Establish Causation	**Straw in the Wind**	**Smoking Gun**
No	*Passing* affirms relevance of hypothesis but does not confirm it. *Failing* suggests hypothesis may not be relevant, but does not eliminate it.	*Passing* confirms hypothesis. *Failing* does not eliminate it.
Yes	**Hoop**	**Doubly Decisive**
and	*Passing* affirms relevance of hypothesis but does not confirm it. *Failing* eliminates it.	*Passing* confirms hypothesis eliminates others. *Failing* eliminates it

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

(b) 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-a` -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 evenhanded 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 c c 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 Germany'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 1989.4 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 1989.5 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

5. Bennett (1999, 2003, 2005). See also English (2000, 2002); Checkel (1997); Gross Stein (1994).
(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 theWest, Brooks's andWohlforth'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 the 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's 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 antireform 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 Strategic Defense Initiative] was being deployed, Eastern Europe was asserting its autonomy, and Soviet clients were losing their counterinsurgency wars in Afghanistan, Angola, and Ethiopia'' (Snyder, 1988: 128).

As it turned out, all these conditions were more than fulfilled within two years, except for the deployment of a working SDI system. Yet apart from the unsuccessful coup attempt of 1990, Soviet hardliners never came close to regaining power. Snyder's theory thus appears to have failed a *hoop test* when the developments he thought would bring the Stalinist coalition back to power indeed took place, but the Stalinists still did not prevail. Conversely, the learning explanation survives a *hoop test* on the basis of evidence that antiinterventionist ideas won out because they resonated with recent Soviet experiences, rather than because their advocates represented a materially powerful coalition.

Despite strong evidence that both material and ideational factors played a role in Soviet restraint in 1989, one variant of the material explanation appears to fail a *hoop test*. Two internal Soviet reports on the situation in Europe in early 1989, one by the International Department (ID) of the Soviet Communist Party and one by the Soviet Institute on the Economy of the World Socialist System (IEMSS in Russian), argued that a crackdown in Eastern Europe would have painful economic consequences for the Soviet Union, including sanctions from the West. The IEMSS report also noted the growing external debts of Soviet allies in Eastern Europe (Bennett 2005: 96–7). At the same time, these reports provide ample evidence for the learning explanation: the IEMSS report warns that a crackdown in Poland could lead to an ''Afghanistan in the Middle of Europe'' (Bennett 2005: 101), and the ID report argues that ''authoritarian methods and direct pressure are clearly obsolete . . . it is very unlikely we would be able to employ the methods of 1956 [the Soviet intervention in Hungary] and 1968 [the Soviet intervention in Czechoslovakia], both as a matter of principle, but also because of unacceptable consequences'' (Bennett 2005: 97).

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.