

- SAGAN, S. D. 1995. *Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press.
- SEKHON, J. S. 2004. Quality meets quantity: case studies, conditional probability and counterfactuals. *Perspectives in Politics*, 2: 281–93.
- SHAFER, M. D. 1988. *Deadly Paradigms: The Failure of U.S. Counterinsurgency Policy*. Princeton, NJ: Princeton University Press.
- SKOCPOL, T. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- and SOMERS, M. 1980. The uses of comparative history in macrosocial inquiry. *Comparative Studies in Society and History*, 22: 147–97.
- STINCHCOMBE, A. L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace.
- SWANK, D. H. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States*. Cambridge: Cambridge University Press.
- TENDLER, J. 1997. *Good Government in the Tropics*. Baltimore: Johns Hopkins University Press.
- TRUMAN, D. B. 1951. *The Governmental Process*. New York: Alfred A. Knopf.
- TSAI, L. 2007. *Accountability without Democracy: How Solidary Groups Provide Public Goods in Rural China*. Cambridge: Cambridge University Press.
- VAN EVERA, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- WAHLKE, J. C. 1979. Pre-behavioralism in political science. *American Political Science Review*, 73: 9–31.
- YASHAR, D. J. 2005. *Contesting Citizenship in Latin America: The Rise of Indigenous Movements and the Postliberal Challenge*. Cambridge: Cambridge University Press.
- YIN, R. K. 2004. *Case Study Anthology*. Thousand Oaks, Calif.: Sage.

## CHAPTER 29

---

INTERVIEWING  
AND QUALITATIVE  
FIELD METHODS:  
PRAGMATISM AND  
PRACTICALITIES

---

BRIAN C. RATHBUN

INTENSIVE interviewing is a powerful, but unfortunately underused tool in political science methodology. Where it is used, it is generally to add a little color to otherwise stiff accounts. Rarely do researchers talk to more than a handful of respondents. There are numerous practical reasons for this. Gaining access to interview subjects, particularly elites, is often difficult. Interviewing is costly as it often entails traveling great distances, sometimes across national borders. Interviewing often requires tremendous personal investment in language training that might not seem worth it. It is often a risky strategy. Even after the hurdles of access and travel are overcome, informants might reveal little. However, these obstacles cannot fully explain why more political scientists do not utilize interviewing in their research as a major source of data, or even as a supplement to quantitative analysis or archival records.

I maintain that there are two reasons why interviewing is often underused. First, interviewing often runs afoul of methodological tendencies in the discipline. Certain precepts of what I call the naive versions of behavioralism and rationalism make many skeptical about interviewing. Naive behavioralism objects to the status of data derived from interviewing as it is by nature subjective and imprecise, and therefore subject to

multiple interpretations. Naive rationalism aims at a framework for understanding politics that privileges structure over agency and often assumes that the political actors of interest view the world objectively and respond the same way to the same stimuli. Both tendencies have the result of privileging some data over others. Most suspect is the kind of information that interviewing is best at ascertaining and that is crucial for many of the most interesting research questions in political science. To the extent that interviewing is endorsed, it is often with some questionable advice and presuppositions: that interview data should always be treated as opinion and not fact, that questions should be indirect and concrete rather than reflective and direct, and that dissembling on the part of respondents is endemic and irremediable. The first section of this article makes the pragmatic case for interviewing. Interviewing, despite its flaws, is often the best tool for establishing how subjective factors influence political decision-making, the motivations of those involved, and the role of agency in events of interest. Behavioralism and rationalism alert us rightly to the importance of rigor in our analyses, but there are steps that can be taken to eliminate some of the concerns about reliability and validity. Skepticism should not be exaggerated.

The second reason why interviewing is often underused is that its practicalities are often not taught. The second portion of this chapter is therefore devoted to assembling in one place the consensus in the literature on the basics of how to undertake interviews, including issues of how to build arguments using interview data, how to structure questionnaires, the proper role to adopt vis-à-vis respondents, and how to gain access to conversation partners. Although advocates propose different techniques for interviewing, the differences are fairly trivial, often just the use of different terms or typologies to describe the same process. There is a striking degree of agreement among proponents once the initial issue of the worth of interviewing data is resolved.

Both tendencies are symbolized perhaps best by the scant attention devoted to interviewing in what is now possibly the most common methodological text in the political discipline, King, Keohane, and Verba's *Designing Social Inquiry* (1994). This book devotes just a single footnote to the issue, a footnote that I argue points in many wrong directions as it is based on the application of the lessons of naive behavioralism to small-*N*, intensive case-study projects, those where interviewing can have the biggest impact. This chapter will focus on what has come to be known as "semi-structured" interviewing as opposed to "open-ended" interviews more common in ethnography or "close-ended" interviews used in surveys (Mishler 1986). Although ethnography and semi-structured interviewing share a commitment to probing in-depth the experiences of respondents and generally stress context over generalizability, induction over deduction, and complexity over parsimony, true ethnography is rare in modern political science for a variety of reasons. Political science generally involves the explanation of particular events. Participant observation, the most significant feature of ethnography, is generally not even possible as the outcome of interest has usually already taken place (although it can certainly be a fruitful source of new research questions and puzzles). Driven by explanation, political science generally involves the testing of hypotheses, even if they develop inductively over the course

of the research process. This requires a more directed research strategy in which researchers seek to uncover some degree of objective truth. The ethnographic style begins with a clean slate and no such presuppositions and is associated with a more relativistic epistemology in which there is no real establishment of fact (Leech 2002, 665). Still there are lessons for ethnographers in the literature on semi-structured interviewing.

## 1 SKEPTICISM ABOUT INTERVIEWING IN POLITICAL SCIENCE

Methodological trends in modern political science, most notably behavioralism and more recently rational choice, have focused our attention on the importance of rigorous analysis based on principles drawn from the natural sciences. Most critical has been the value of objectivity and theory-driven research. Early behavioralists stressed the importance of freeing explanation from the normative values of the researcher. Data should also be as objective as possible, so that different researchers could agree on the meaning of the same piece of data, the social scientific equivalent of replication. The explanation of political events should be pursued separately from normative theory. And lest personal values should nevertheless influence our results, pure research was given increasing priority over applied research. Quantification was not pursued for its own sake, but rather would improve the precision of data and facilitate objective evaluation. Numbers are a more universal language less prone to differences in interpretation (Easton 1962; Somit and Tanenhaus 1967).

This led naturally to a tendency to focus on observed behavior, rather than unobservables such as mental processes that were more prone to subjective evaluation. Behavioralism of course did not completely eschew studies of political phenomena without observable behavioral implications. It did not go as far as the behaviorists in economics and psychology (Merkl 1969). Nor did it abandon the examination of the effects of the subjective perceptions of political actors on political events. The advances made in public opinion research are an example of both. However, whenever feasible, the possible subjectivity of political actors was to be made more manageable, for instance by the deductive designation of a limited number of response categories in a survey so as to allow objective evaluation, precision, and comparison by the researcher. Behavioralists prefer "structured" interviews, in which the same questions are asked in the same order with a restricted number of answers (Patton 1990, 277-90). This of course also served the goal of capturing the generalizability of political phenomena. Intensive interviewing was simply too costly and time consuming if a survey researcher wanted to gain an overall perspective. A more in-depth approach could be used, however, in early stages to make sure surveys were appropriately designed with questions that adequately measured variables of key concern.

Research was also to be driven by theory. This goal was juxtaposed unfavorably against traditional political science, which was marked by description. Without theory, political science is just journalism (and with a smaller readership). Theories were also to be of a particular sort—generalizable and parsimonious. Scholars should try to explain a lot with a little (Easton 1962; Somit and Tanenhaus 1967). The turn towards generalization was generally associated with a focus on structure over agency. The more that human beings had choice and could change their environment, the less generalizable or parsimonious theories could be (Almond and Genco 1977). Traditionalists responded that politics was far too complex and rich to be captured by general theories. Context mattered tremendously. Behavioralists replied that they missed the forest for the trees. A theoretical rather than an empiricist focus also meant that theories were best deduced from logical principles, rather than inductively built from the ground up with empirical data.

Interviewing does not always stand up well to these standards, as fruitful as they have been for modern political science. While behavioralism is well suited to large-*N* survey research, its principles have increasingly been applied in naive form to small-*N* case-study research. Concerns about the scientific status of data drawn from interviews are rarely explicitly stated, but rather are more subtle. They are expressed in advice from faculty advisers to their graduate students or in skepticism about interview data in peer reviews of work submitted to academic journals. Interviewing's faults lie in its inherent emphasis on complexity and context to the detriment of objectivity, parsimony, and generalizability. The very purpose of interviewing is generally to go in-depth in a way that secondhand sources, archives, or surveys do not allow (Berry 2002, 682). Interview subjects are often chosen because of their unique perspective on a particular phenomenon or event. Interviewing is of most use when interviewees have shaped the world around them, often undermining the goal of generalizability. Questions are not standardized across respondents, impeding comparability. The data, generally in the form of quotes or statements by the respondents, are prone to multiple, subjective interpretations by the researcher, making them less reliable. As a result, information culled from interviews has a dubious status, often described merely as the opinion of the respondent rather than fact. Because it cannot be objectively verified, it is less trustworthy. In any case, it is of little use in building the broader arguments that behavioralism is interested in. Interviewing is often regarded as the method used when social scientists do not have fully formed ideas or theories, leading to critiques of inductivism (Mishler 1986, 28).

The usefulness of interviewing has also come increasingly into question with the growing popularity of formal and rational choice frameworks for studying politics. Behavioralists did not deny the importance of unobservables such as motivations, psychology, norms, culture, learning, ethics, all of which might fall under the broader rubric of ideational factors. They merely prefer more observable factors because they lend themselves to more objective analysis akin to the natural sciences. These other factors are not denied, but ignored (Almond and Genco 1977). Rationalists, in contrast, make particular ontological assumptions in their efforts to build general models of politics, most importantly the objectivity of political actors. While they might have

varying preferences, all actors placed in the same situation view stimuli similarly, and if they share the same preferences, will respond identically (Mercer 1996; Parsons 2006). The importance of objectivity is different than in behavioralism. For the latter, it is necessary for the researcher. For rationalists, it is necessary for political actors as well. If the situational incentives are the key to explanation, agency is limited, and perception is unproblematic, there is little need for interviewing. Rationalists "eliminate the mind" from analysis (Mercer 2005). They restrict their ontology for the broader methodological goal of generalizable theory (Fearon and Wendt 2002; Almond and Genco 1977). Given that interviewing is often used for deciphering what goes on between the ears, it serves less of a purpose for them. It might be of use for establishing preferences, but rationalists generally prefer the deduction of preferences from broader theories. They adopt the behavioral line about establishing motivations, as they are unobservable (Frieden 1999).

The rising influence of rational choice theory also has a more subtle but perhaps even more important effect. Conceptualizing most political situations as games of strategic interaction has increasingly alerted political scientists to concerns about strategic reconstruction by respondents. Interview data are inherently faulty because respondents have incentives to dissemble. This raises the validity issue in interviewing (Brenner 1985; Berry 2002). Respondents might seek to preserve their reputation and legacy or retain private information in an ongoing bargaining situation. When combined with tendencies discovered by psychologists to want to be perceived in a favorable light or justify one's actions to oneself, there are additional grounds not to trust interview data, not because of its epistemological status as in behavioralism, but because interviewees cannot be trusted. As argued below, this is a particularly vexing problem for students of political science. Politics involves power, which introduces unique dynamics. Political science is interested more than the other social sciences in explaining events, often those with broad social impact. In this sense it has more in common with history. For powerful individuals who often operate in an accountable and public realm, this introduces questions of how accounts are framed. Those formerly involved in politics are concerned about their legacy, those still involved about the public perception of their work. As such, political scientists must be particularly attuned to the issue of strategic reconstruction of events to suit the more vested interests of interviewees.

Paradoxically, the proponents of interviewing add fuel to this skeptical fire. There is a strange meeting of the minds between positivists and relativists. Many books on the method stress that it is not for the researcher to judge the correct view of events. Key proponents of the method discuss the difficulty of a true understanding between interviewer and interviewee because each unwittingly brings assumptions and biases to the conversation (Mishler 1986; Rubin and Rubin 1995). Interviewee data are inherently subjective. This is undoubtedly true, but at the extreme, explanation, however difficult and incomplete, ceases to become the goal of social science. This tendency becomes more pronounced in the literature on ethnography as opposed to semi-structured interviewing. When the research question is the attitudes or perspectives on politics of a particular group, the lack of an evidentiary standard might

be appropriate. All are equally entitled to their opinion. But more often political scientists are interested in explaining outcomes, and must make judgements about the accuracy of conflicting accounts. Ironically, relativists end up sounding like positivists when they stress that interview data are more opinion than fact, although they reach that conclusion from different angles. Positivists treat interview data skeptically because objective interpretation is difficult among researchers, relativists because interpretation is an inappropriate imposition of the researcher's subjective views and does not capture the true meaning of the responses. Rationalists assume complete objectivity among the political actors they study. Relativists do the opposite, presuming complete subjectivity to the degree that explanation becomes impossible.

## 2 THE PRAGMATIC CASE FOR INTERVIEWING

Interviewers must be careful to separate fact from opinion, guard against phoney testimony by their conversation partners, should not let their own theoretical or political beliefs affect their interpretation, and cannot be satisfied with mere impressions. These are all valid concerns. But behavioralism and rationalism, taken to an extreme, can have the pernicious effect of discarding potentially valuable (and often invaluable) information. These rules limit the potential of interviewing significantly. Interviewing has flaws, but on pragmatic grounds, it is often the only means to obtain particular kinds of information. As discussed above, jettisoning interviewing often means restricting the set of independent variables to observable, structural factors. This impoverishes the study of politics to an unacceptable degree. Interviewing is often the best-suited method for establishing the importance of agency or ideational factors such as culture, norms, ethics, perception, learning, and cognition (Aberbach and Rockman 2002, 673; Mishler 1986, 279; Rubin and Rubin 1995, 19). Interviewing is often the most productive approach when influence over the outcome of interest was restricted to a few select decision-makers, creating a bottleneck of political power that increases the importance of agency in the story (and also makes interviewing less costly and time consuming). In short, interviewing, whatever its flaws, is often the best-suited method for gathering data on those characteristics of the social world that differentiate it from the natural world: human beings' effort to intentionally transform their environment on the basis of cognition, reflection, and learning (Almond and Genco 1977).

Interviewing is often necessary for establishing motivations and preferences. Even for those approaches that focus on situational incentives, this is an absolutely critical element of any theory. Without an understanding of desires, even the most rigorous rationalist argument will not be falsifiable if it simply infers preferences from observable behavior and a posited set of situational constraints. It falls into the same trap as folk psychology (Mercer 2005). Desires and beliefs must be measured independent of

action. In some instances, interviewing is the only way to do so. For all its problems, interviewing is often more "scientific" than other methods.

Even while interviewing is well suited for discovering preferences and agency, this is not to say that it is not useful at establishing structural causes as well. Interviewing can help establish whether a political actor felt under pressure from forces beyond his or her control, and what those forces were, particularly when there are multiple independent variables in the theoretical mix. To the extent that one is interested, like rationalists, in showing the importance of structural or situational factors in explaining behavior, interviewing can help build the appropriate model. Strategic circumstances might be found by a model to provide the best account for particular action, but are only empirically useful if the model reflects more or less the actual circumstances that decision-makers found themselves in. Interviewing is one, although certainly not the only, way of identifying that situation.

Other methods might provide some insight into these factors, such as archives or memoirs. But interviewing is unique in that it allows the interviewer to ask the questions that he or she wants answered. Memoirs and secondary accounts force the researcher to answer his or her key questions based on what others wanted to write about. Archives sometimes provide insights into the thinking of key individuals, but motivations and agency must be more indirectly established than in interviewing. They require considerable inference and interpretation, raising the reliability issue to a greater degree than in interviewing. Interviewing has the advantage of being perhaps the most directed and targeted method in the qualitative arsenal.

To the extent that interviewing is endorsed by many in the behavioralist and rationalist tradition, it comes with important provisos that are often ill advised and overly constricting. A single passage, taken from a footnote in Keohane, King, and Verba's guide to qualitative research (now standard reading in methodology classes in political science graduate courses), contains many of these:

Be as concrete as possible . . . In general and wherever possible, we must not ask an interviewee to do our work for us. It is best not to ask for estimates of causal effects. We must ask for measures of the explanatory and dependent variables, and estimate the causal effects ourselves. We must not ask for motivations, but rather for facts. This rule is not meant to imply that we should never ask people why they did something. Indeed, asking about motivations is often a productive means of generating hypotheses. Self-reported motivations may also be a useful set of observable implications. However, the answers must be interpreted as the interviewee's response to the researcher's question, not necessarily as the correct answer. If questions such as these are to be of use, we should design research so that a particular answer given (with whatever justification, embellishments, lies or selective memories we may encounter) is an observable implication. (1994, 112 n.)

This quote raises four issues: the extent to which interview data have the status of fact; how tangible questions should be; whether data should be pursued directly; and whether respondents can be trusted to give accurate accounts.

According to these scholars, responses from interviewees cannot be taken as "the correct answer," i.e. data with the status of fact, but rather should be treated as

opinions. Motivations are juxtaposed antithetically to facts, indicating a bias in favor of behavior. We should ask for concrete actions, what an interviewee did, rather than his or her overall impression of a situation. But behavior rarely speaks for itself. There are often multiple possible causes and motivations for the same action and outcome. Research is often driven by the desire to explain an already determined and known behavioral outcome, likely due to its intrinsic importance. This means that social scientists will be trying to gauge the effect of competing independent variables with the same expected effect. In these cases behavior does not speak for itself, and the meaning of behavior must be established. There is no substitute for a good research design, but interviewing might often be the only means for establishing causation in these cases, particularly when the phenomenon of study is relatively recent and there are no reliable alternative sources for judging motivation. Although Keohane, King, and Verba (1994) caution us not to ask respondents to estimate causal effects or state their motivations, this seems unwise. Interviews often involve conversations with individuals in a unique position to gauge the importance of multiple and equally plausible causal factors, which any research question of interest generally suffers from. We need not instantly accept those estimates as fact, but when a consensus appears among those in a best position to know, it should be taken very seriously. Presuming that each account is equally valid, as many advocates of interviewing suggest, goes too far, however. Social science requires the researcher to weigh conflicting evidence and offer the best interpretation possible supported by the evidence. Understanding who was in a position to know the true facts of the outcome the researcher is trying to explain is a key factor in balancing different accounts, as it helps separate opinion from fact.

When Keohane et al. recommend concreteness, they mean that questions should be both tangible and indirect. Others do as well (Mishler 1986, 315). In terms of the former, more grounded questions about facts are less subject to multiple interpretations. Answers to concrete questions, such as the temporal sequence of particular actions, are often useful in reconstructing events and mediating between competing arguments. But questions that ask an interviewee to reflect more abstractly about the underlying causes and motivations behind his actions are just as important. In many instances, the researcher already knows about behavior because it has been a matter of public interest and record that led to the selection of the topic in the first place. It is the motivation that is missing. In practical terms, answers to less concrete questions are often easier to obtain, particularly as considerable time may have passed between the interview and the events or outcome in question. Contemporary newspaper accounts or even archives can often provide the facts about what happened on any particular day. To the extent that this is true, interviews should be devoted to allowing respondents to reflect on their experience, which they might only now have had an opportunity to do. Particularly when dealing with elites, respondents are perfectly able to contemplate the broader meaning of their actions just as well as a political scientist, although they do so in less self-consciously theoretical terms. One revealing technique is to ask conversation partners to informally test competing arguments against one another. Rarely are political science theories too complicated to explain

to an intelligent layman. The researcher can pose the question as, "Some say X, others say Y, how would you respond to that?" (Mishler 1986, 318). By posing such trade-offs, researchers might encourage interviewees to draw evidence to support their account. This is the best of both worlds, for respondents to explain the broader significance of their actions with actual data about what they did. Another useful method is to ask counterfactuals in the form of "why didn't you do X" or "what if you had done Y in situation Z?" This is a very useful technique for establishing the extent to which structure or agency prevailed. The response might be that one path was more valued than another, or that one path was blocked by some constraint.

Being concrete also means questions should be indirect (McCracken 1988, 21; Brenner 1985, 151). Keohane et al. (1994) recommend for instance not to ask a white conservative whether he is racist, but rather if he would mind if his daughter married a black man (1994, 112 n.). Indirect questions are particularly necessary in instances of what are known as "valence" issues, in which there is only one publicly acceptable position to have on an issue. Any outcome of interest in which politicians might have sought narrower parochial interests at the expense of the broader societal good, which is much of political science, might be so loaded and is a key place to worry about strategic reconstruction. On issues in which there are multiple, publicly acceptable positions, this is less of an issue. Skeptics of interviewing also worry that directed questions put words in the mouth of respondents and are therefore inappropriate. This is a potential worry, and researchers should indicate in their research, perhaps in a footnote, the question to which a quote corresponds if there is a fear of leading respondents.

However, as in journalism, interviewing's payoff is often in the quotation, getting someone on the record, even if it is nonattributable (Mishler 1986, 347). That is, direct statements are more valuable in terms of impact and credibility. And one must remember that interviews with subjects in political science often do have a strategic component, and interviewers must sometimes lead respondents into answering rather than dodging their questions, while simultaneously avoiding giving them the answers. Interviewers should proceed indirectly so as to get a sense of how far respondents will go, but ask follow-up questions, trying to establish more directly what one wants to know. If the interviewee goes so far, one can then safely ask for a confirmation in the form of "I hear you saying that . . .," without fears of leading the witness, as it were. Questions cannot be purely open ended and undirected. A couple of techniques might be helpful. In cases in which there is a publicly stated position for a particular behavioral outcome but the researcher or others have a suspicion that there was an ulterior motive not flattering to the interviewee, researchers can use presupposition questions in which they indicate their acceptance of the ulterior motive as perfectly natural, assumed, and justifiable (Leech 2002, 666; Mishler 1986, 304). Interviewers might even note that their actions were perfectly understandable in light of the circumstances. They might use euphemisms that take the hard edges off that behavior (Leech 2002, 666). Or they might mention (only honestly!) that other interview subjects have indicated as much to them already. They should then look to see if the subject acquiesces or seeks to correct the assumption. The former

is revealing if not concrete proof; the latter provides an opportunity for the interviewer to challenge the account and ask the conversation partner to provide evidence for it.

Although the concern about intentionally (or unintentionally) inaccurate answers is real, it is often overstated (Rubin and Rubin 1995, 225; Berry 2002, 688). When Keohane et al. refer to "justification, embellishments, lies and selective memories," there is almost a presumption of disingenuousness on the part of the respondent. But my experience with interviewing is that dishonesty is not the norm, even among the most high-ranking officials. I was able to solicit evidence on recent issues in national security policy that did not cast subjects in a positive light (Rathbun 2004). Most respondents would rather refuse an appointment than spend the time averting a difficult subject. Interviewees are more likely to commit sins of omission than commission, avoiding deliberate falsehoods and attempting to steer the conversation to other aspects of the subject. For someone who has prepared properly, this becomes obvious. Researchers are not nearly as threatening as journalists. An academic's audience is more limited and is certainly not the average man on the street reading tomorrow's newspaper. This creates difficulty in gaining access to important respondents, but once access is gained, they might be more forthcoming. Interviewees might welcome the chance to explain their actions more fully with fewer fears of being boxed in and quoted out of context (Rubin and Rubin 1995, 105). Those engaged in politics often believe in what they do and are proud of their accomplishments, even when they seem distasteful to some. Certainly the risk of being lied to is surely outweighed by the potential of new and exclusive data. Researchers should not presume interview subjects are telling the truth, but they should not assume they are lying either.

There are a number of other ways of avoiding the perils of strategic reconstruction. Some come from the common sense of journalism. Most simply, when possible, find multiple sources for the same data, either written or interview based (Berry 2002, 680; Rubin and Rubin 1995, 225-7; Brenner 1985, 155). If this is done with other interviews, this reinforces one's belief that what one is hearing is more than just opinion or that it is just a particular interpretation of ambiguous information. And it helps reduce worries that some respondents exaggerate the role they played in the process (Berry 2002, 680). Let them know you are an expert by your questioning, even if it means the gratuitous use of jargon. This tells them you are an informed observer who cannot be easily manipulated (Rubin and Rubin 1995, 197). Confront the respondent with data contrary to his or her point of view, even if you might believe he or she is being straightforward, as this might provide ideas about how to get leverage on multiple causal factors (Rubin and Rubin 1995, 223). The fact that much of the study of politics revolves around conflict in the public sphere also has advantages in gaining more accurate accounts. Interviewees can be challenged with data drawn from the other side of the political spectrum, a standard journalistic technique (Rubin and Rubin 1995, 68-70, 217). This makes the interviewee appear more neutral than if the challenge is issued directly, although interviewers should be careful not to give the impression that they are engaging in the same "gotcha" game as the media. If such an opponent is unavailable, researchers can explicitly refer themselves as a "devil's

advocate" in order to be less provocative (Rubin and Rubin 1995, 215). Respondents might be asked to critique their own case, to explain why their adversaries do not accept their account (Berry 2002, 680).

When bias is expected, one can pose some preliminary questions unrelated to the subject that gauge a general ideological approach and allow the researcher to discount certain aspects of testimony, sometimes called "norming" (Rubin and Rubin 1995, 210). Understand the particular bureaucratic, political, or ideological position of your interviewees so you know what biases to expect. Know what the party line might be (Rubin and Rubin 1995, 221). This requires a familiarity with the public record. In improvisational questions, probe differences between the public account and the one you are hearing, but do not despair if they vary. On the contrary, generally this means that you are uncovering something new and exciting. When a researcher finds differences it can yield important new insights about the political context that respondents were negotiating. Follow up on this, as it is important for gauging the significance of structure and agency. Ask why particular arguments were instrumental and why this was necessary. If memories appear weak, ask basic questions about details that allow one to assess the overall credibility of the source (Rubin and Rubin 1995, 84). A researcher might make deliberate mistakes that should be corrected if memories are good (Rubin and Rubin 1995, 219). Finally, if the account has internal contradictions or inconsistencies, let these be known to the respondent and note the adjustments that he or she makes to the story.

### 3 THE CONSENSUS ON THE PRACTICALITIES OF INTERVIEWING

While there are often disagreements about whether interviews are worthwhile ventures, among those who undertake them there is a general consensus on most points. This section reviews those major points of agreement, proceeding chronologically from preparation to interpretation and writing.

Perhaps the most important lesson for any would-be interviewer is not to begin intensive interviewing without being fully prepared (Brenner 1985, 152). As McCracken writes, "Every qualitative interview is, potentially, the victim of a shapeless inquiry. The scholar who does not control these data will surely sink without a trace" (1988, 22). The interviewer should exhaust all of the secondary sources and publicly available primary sources before beginning (Berry 2002, 680). This allows him or her to work out a puzzle, if one isn't already evident; to figure out what is known and what is not known so questions can be more targeted and efficient; to understand how the debate is framed in different contexts (nation states, cultures, time periods, etc.) in the case that the issue involves political conflict, as it generally does in political

science; and to develop expectations about what interview data would be evidence for his or her initial hypothesis or other competitors. In those instances in which research is being undertaken in a new area without a significant paper trail, a political scientist might consider a set of exploratory interviews to get a better sense of the interesting theoretical issues (McCracken 1988, 48), but these should be limited to lower-ranking respondents or those with knowledge of a phenomenon but not directly involved, such as journalists. The researcher should not show up with a year's fellowship in a new place hoping that things will fall into place on the basis of ongoing conversations. They might, but luck will be necessary. An understanding of the literature will not only help the researcher eventually frame his argument in a broader context in the opening passages of an academic work but also provide a benchmark for what is surprising in the interview data he collects (McCracken 1988, 31). A good scholar will read not only the qualitative work in his field but also the quantitative even if he is not a statistical whiz, as it is a source of hypotheses and sometimes an indication of what is yet to be explored due to the inherent limitations of quantitative analyses, such as difficulties in establishing causation.

It is crucial to remember that the literature review, or one's own hypothesis, must be held at arm's length, so that researchers are open, in the true scientific sense, to other possibilities. McCracken calls this "manufacturing distance" (1988, 23). This is particularly important in interviewing as the method is often used when there is a black hole in the evidence about something of theoretical interest, in which case assumptions would be inappropriate. And whereas quantitative analyses might be rerun or a secondary source reread, interviewers often only have one chance with their sources of data. If you are insistent on a particular hypothesis, even if it is not working out, you will waste an experience that cannot be repeated. "A good literature review makes the investigator the master, not the captive, of previous scholarship" (McCracken 1988, 31). Previous theories, or your own, might act as blinders (Rubin and Rubin 1995, 64). This is not only good social science, but any possibility of overturning conventional wisdom benefits the scholar's career!

After the literature review, researchers should establish contact with those who have a particular knowledge of the outcomes of interest. Exploratory interviews are often fine sources of such contacts. It is generally best to begin with low-ranking members of the group you are interested in, whether it be a trade union, government bureaucracy, or social movement. The researcher should talk to anyone who will see him or her. This allows the social scientist to hone and try out different framing of questions, develop and accumulate contacts (sometimes known as "snowballing"), and gather basic information so as to be better prepared for more senior respondents who have less time and tolerance. Initial contact should not be a cold call if it can be avoided. Many recommend a professional letter (Goldstein 2002, 671; Aberbach and Rockman 2002, 674; Peabody et al. 1990, 453).<sup>1</sup> Letterhead conveys a sense of professionalism and status. In closed societies, it is often good to have a local affiliation at a

<sup>1</sup> Some note that in developing societies requests for interviews should be made in person if possible, for both cultural and technical reasons. See Rivera, Kozyreva, and Sarovskii (2002).

research center. In more open societies, your own foreign university emblem might be better as it helps you stand out from the flood of mail that many of our conversation partners receive. Explain why the respondent might be of particular help for your project. This provides background but also gives them a sense of relevance that is inevitably flattering (Rubin and Rubin 1995, 103). Tell them they are not obligated to speak on the record or on an attributable basis. When possible, mention the person who referred you and others you have talked to, so as to give yourself a sense of credibility. Preferably the former should be on the same side of any political conflict as your contact, but often the latter can be someone opposed, as this might induce an interest on their part in equal time and setting the record straight that might make them more likely to meet you. Follow up this initial contact with a phone call to the individual or his or her appointments assistant. When possible, allow a large timeframe to set up a meeting, making it harder for the contact to refuse you and enabling the inevitable rescheduling. If possible, have a direct phone number, even a cell phone (Goldstein 2002, 671). Playing phone tag from your hotel room voice mail will prove difficult. Gather the names of the administrative staff as you set up the interview, keep them in your mind as you travel to the interview, and seek them out as you enter or leave the office to thank them directly by name.

Most advocates of interviewing recommend recording interviews (Aberbach and Rockman 2002, 674). Tell your interviewee that you would like to record the interview, and if it might make your work more credible later to identify him or her by name, that you would like for it to be on the record. If this is not possible, inform him or her of your preference for it to be nonattributable, in which you might quote, but not directly identify his or her name. Although recording on the record is often thought to make respondents more circumspect, this presumption is based on an assumption that respondents will not be forthcoming in the first place. As I think this is often overstated, I believe the advantages likely outweigh the disadvantages although every circumstance is different, and they obviously vary by the stakes involved and the local culture. "Senior official" does not have the same cachet in print as "Prime Minister Smith." Having a recorded version of the conversation allows one to think through potential follow-up questions when the conversation takes an interesting turn without worrying about taking down the exact text (Brenner 1985, 154; Berry 2002, 682; Douglas 1985, 83; McCracken 1988, 41). Such a recording allows you to assemble a verbatim transcript that you can make available in an effort to remedy some of the subjective interpretation problems endemic to the method (Brenner 1985, 155; Mishler 1986, 36; Rubin and Rubin 1995, 85). This transparency helps avoid the accusation of massaging and spinning the data. Given that your research question or hypothesis will likely shift as you accumulate information, such a transcript is often invaluable as you have information on record that you might not have thought relevant and noted before in your notes. And given that interviewing is often used to identify the motivation and meaning behind otherwise indeterminate behavior, specific quotations are sometimes the only evidence possible or even moderately convincing. Merely stating your hypothesis and providing a footnote of "Interview with Mr X" is not sufficient. However, allow the interviewee the option of going off

the record or not for attribution at any particular point in the interview. They often take it. Make sure you are clear on what you or they mean as "off the record" or "on background" (Goldstein 2002, 671). And take notes so that you do not have to transcribe every single recording, saving that valuable time for particularly evocative conversations.

To the degree possible, the social scientist should inform himself of the role played by that individual in the phenomenon of interest. Let that preparation be known in a brief introduction of yourself and your general research interests. This helps break the ice (Leech 2002, 666; Peabody et al. 1990, 652). Yet the researcher does better not to specifically identify his or her own working hypothesis, whether it be a fully formed argument or just a hunch, as the respondent might be tempted inappropriately to help gather information on behalf of the interviewer or presumptuously dismiss it (and you) (McCracken 1988, 27–8; Aberbach and Rockman 2002, 673). In this sense, Keohane et al. are correct that interviewees should not do our work for us. Use specialized language or acronyms to demonstrate your competence and prompt them to avoid needless background that wastes valuable time (Rubin and Rubin 1995, 77–9). When possible, interview in the language of the interviewee. All of this demonstrates professionalism and gives the sense that the conversation partner is truly valued and important. While politicians might be accustomed to this treatment, other (even powerful) figures might not. The social scientist must strike a balance between formality on the one hand and rapport on the other. Professionalism and a certain distance ensure that the interviewer retains control of the questioning and is taken seriously. Yet too cold a demeanor can inhibit respondents from trusting the interviewer with often sensitive information (McCracken 1988, 26–7). Questioning should be sympathetic, respectful, and nonargumentative, although this does not imply unchallenging, as discussed earlier (Brenner 1985, 157). Yet political science interviewing differs from journalism in that it must be less adversarial. Interviewees might have an obligation to the public interest to sit for an interview with a journalist and they can hardly call it off midstream. That is not the case with academics. Some suggest that an academic demonstrate less of a command of the subject matter than he or she actually has so that respondents do not feel inferior, but this runs the risk of making some feel they are wasting their time (McCracken 1988). Leech's advice to appear knowledgeable, but not as knowledgeable as the interviewee, is probably better (2002, 665).

Researchers should bring a questionnaire to their interviews specifically designed around the unique role played by their particular conversation partner. This helps separate fact from opinion as well as not wasting time with information that could be gathered from other sources (Leech 2002, 666). Interviews should not be entirely improvisational because researchers might forget important questions and not have an opportunity to fill those holes later. But they cannot be too rigidly designed so as to inhibit adaptation. After all, the goal of interviews is rarely simply confirmatory. Social scientists want to be surprised. They are often interviewing in the first place because there are certain things that are just not known. The questionnaire should be a general set of themes, and specifically worded questions that have been found

to elicit better responses than others. Do not begin with challenging questions as they might put the respondent on guard and heighten sensitivity to the running tape recorder (McCracken 1988, 38; Mishler 1986, 294; Douglas 1985, 138; Rubin and Rubin 1995, 223). They should be reserved until later after a sense of rapport has been established (but don't wait too long and run out of time) (Leech 2002, 666). But beginning questions should not be tedious either. The interviewer might first try to elicit facts and then as the interviewee loosens up begin with more complex matters of interpretation, motivation, etc., particularly as the former often jogs the memory of events long past (Peabody et al. 1990, 452; Rubin and Rubin 1995, 210). If time is limited, however, focus first on the absolute essentials and then work towards other less important matters (Rubin and Rubin 1995, 200). In these instances, do not ask questions for which information can be obtained from other less busy sources or that you should know the answer to already, thus undermining your credibility (Peabody et al. 1990, 452). Learn the value of nonverbal communication. If an answer is not complete, the researcher can indicate this by remaining silent and nodding his head, both signaling an interest in further explication. If a respondent avoids a question or simply misses it, note this and return to the subject later from a different angle. Have at your disposal some "bridges" that enable you to return to the area of interest in a respectful way (Berry 2002, 681).

When leaving, always ask the conversation partner for other suggestions about whom to talk to. Those directly involved are often in the best position to know who would be uniquely knowledgeable on some aspect of your research topic and can be used as a reference in your request. One should never finish an interview without knowing the next person one wants to talk to. An interviewer should never have crossed off all the names of possible subjects. Also, always ask if he or she would be available for additional questions, either in person, e-mail, or by phone. This is a promise that can be included in a follow-up letter requesting more information that has a better chance of getting past the secretary, especially if you have been previously cordial in your interactions (Rubin and Rubin 1995, 138).

Successive interviews should build on each other. The researcher should constantly be looking over his notes, looking for sources of agreement and disagreement, new themes and causal factors previously untheorized or not mentioned in the literature. Over time, interviews will naturally become less exploratory or inductive and more deductive (Rubin and Rubin 1995, 43; Brenner 1985, 152; McCracken 1988, 45). Scholars will begin to develop more precise hypotheses that they seek to confirm or test the limits of. Questions will become more honed and precise and fewer topics will need to be explored as a narrative emerges. As with all research there will be a point of diminishing returns in which later interviews generate less new information than previous ones, although interviewers should be aware of premature closure (McCracken 1988, 45). This will guide the social scientist in his effort at organizing notes and thoughts once interviewing is complete. With the proper legwork before interviewing and continual reappraisals of the data during the process, researchers can end the data collection phase of research with a well-formed albeit sometimes preliminary idea of the final product.



## 4 CONCLUSION

Interviewing is a potentially invaluable tool for collecting data that researchers should approach pragmatically, both in terms of how, but also whether, it is conducted. Its financial costs and methodological difficulties are real, but if a political scientist believes there might be some aspect of a phenomenon best captured by directly talking to participants, he or she should not hesitate. Doubts about the status of interview data and the reliability of respondents must be taken into account but can be addressed. These disadvantages rarely outweigh the unique advantages of interviewing: the ability to target questions directly to actual participants and push them for responses in a way that archival or other qualitative research never allows. Although most would agree that interviewing is a kind of art that takes some practice, there is a solid consensus on the practicalities of interviewing. It is no substitute for careful research and investigation of secondary and written primary sources, however. Systematic analysis and meticulous preparation are a must in any kind of analysis.

## REFERENCES

- ABERBACH, J. D., and ROCKMAN, B. A. 2002. Conducting and coding elite interviews. *Political Science and Politics*, 35: 673-4.
- ALMOND, G. A., and GENCO, S. J. 1977. Clocks, clouds and the study of world politics. *World Politics*, 29: 489-522.
- BERRY, J. M. 2002. Validity and reliability issues in elite interviewing. *Political Science and Politics*, 35: 679-82.
- BRENNER, M. 1985. Intensive interviewing. Pp. 147-62 in *The Research Interview: Uses and Approaches*, ed. M. Brenner, J. Brown, and D. Canter. London: Academic Press.
- DOUGLAS, J. D. 1985. *Creative Interviewing*. Beverly Hills, Calif.: Sage.
- EASTON, D. 1962. The current meaning of "behavioralism" in political science. Pp. 1-25 in *The Limits of Behavioralism in Political Science*, ed. J. C. Charlesworth. Philadelphia: American Academy of Political and Social Science.
- FEARON, J., and WENDT, A. 2002. Rationalism v. constructivism: a skeptical view. In *Handbook of International Relations*, ed. W. Carlsnaes, T. Risse, and B. Simmons. London: Sage.
- FRIEDEN, J. A. 1999. Actors and preferences in international relations. Pp. 39-78 in *Strategic Choice and International Relations*, ed. D. A. Lake, and R. Powell. Princeton, NJ: Princeton University Press.
- GOLDSTEIN, K. 2002. Getting in the door: sampling and completing elite interviews. *Political Science and Politics*, 35: 669-72.
- KING, G., KEOHANE, R., and VERBA, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- LEECH, B. L. 2002. Asking questions: techniques for semistructured interviews. *Political Science and Politics*, 35: 665-8.
- MCCRACKEN, G. 1988. *The Long Interview*. Qualitative Research Methods, 13. Newbury Park, Calif.: Sage.

- MERCER, J. 1996. *Reputation and International Politics*. Ithaca, NY: Cornell University Press.
- . 2005. Rationality and psychology in international politics. *International Organization*, 59: 77-106.
- MERKL, P. H. 1969. "Behavioristic" tendencies in American political science. Pp. 141-52 in *Behavioralism in Political Science*, ed. H. Elau. New York: Atherton Press.
- MISHLER, E. 1986. *Research Interviewing: Context and Narrative*. Cambridge, Mass.: Harvard University Press.
- PARSONS, C. 2006. *The Logics of Interpretation*. Oxford: Oxford University Press.
- PATTON, M. Q. 1990. *Qualitative Evaluation and Research Methods*. Newbury Park, Calif.: Sage.
- PEABODY, R. L., HAMMOND, S. W., TORCOM, J., BROWN, L. P., THOMPSON, C., and KOLODYNY, R. 1990. Interviewing political elites. *Political Science and Politics*, 23: 451-5.
- RATHBUN, B. 2004. *Partisan Interventions: European Party Politics and Peace Enforcement in the Balkans*. Ithaca, NY: Cornell University Press.
- RIVERA, S. W., KOZYREVA, P. M., and SAROVSKII, E. G. 2002. Interviewing political elites: lessons from Russia. *Political Science and Politics*, 35: 683-7.
- RUBIN, H. J., and RUBIN, I. S. 1995. *Qualitative Interviewing: The Art of Hearing Data*. Thousand Oaks, Calif.: Sage.
- SOMIT, A., and TANENHAUS, J. 1967. *The Development of American Political Science: From Burgess to Behaviorism*. Boston: Allyn and Bacon.

## CHAPTER 30

---

# PROCESS TRACING: A BAYESIAN PERSPECTIVE

---

ANDREW BENNETT

How are we to judge competing explanatory claims about historical cases? How can we make inferences about which alternative explanations are more convincing, in what ways, and to what degree? How can we make inferences from the explanation of individual historical cases to the general explanatory power and scope conditions of the underlying theories that explanations of cases draw upon? Bayesian logic and case-study methods, particularly methods of within-case analysis such as process tracing, offer two approaches to these questions. Scholars have long recognized that Bayesian analysis is suited to assessing explanations for one or a few cases and for making inferences from the evidence in individual cases to the general viability of theories, and within-case analysis focuses on the historical explanation of individual cases. Recently, experts on qualitative methods have begun to argue that these approaches have much in common (McKeown 2003; Brady and Collier 2004; George and Bennett 2005). The relationship between the two has not been analyzed in any detail, however, nor has there been an analysis of how discussions of Bayesian inference help illuminate the strengths and limits of process tracing. The present chapter seeks to fill these gaps.

This chapter first provides an overview of process tracing, focusing on the dimensions of this method that are relevant to Bayesian logic (for a more complete discussion of process tracing, see George and Bennett 2005). It then briefly summarizes the logic of Bayesian inference, emphasizing the parallels with the logic of process tracing. The chapter illustrates these points with examples from the historical explanation of

political events, including the 1898 Fashoda crisis, the end of the First World War, and the end of the cold war. The chapter concludes that Bayesianism helps us understand the strengths of case-study methods, including their potential to develop and test explanations even with limited evidence drawn from one or a few cases, and their limits, including the provisional nature of historical explanations and the challenges of generalizing from small numbers of cases.

---

## 1 HISTORICAL EXPLANATION: THREE ILLUSTRATIVE CASES

---

The social sciences are rife with debates over the historical explanation of particular cases. Consider three historically and theoretically important examples from my own field of international relations, each of which I return to below. In the first, Kenneth Schultz argues that Britain and France were able to avoid war during the Fashoda crisis of 1898 because democratic processes in Britain reinforced the credibility of its coercive diplomatic demands, while those in France undermined the credibility of its willingness to fight. Others attribute the absence of a war over the Fashoda crisis to differences in French and British military power, or to the democratic values held by the two countries (Schultz 2001). In a second example, Hein Goemans argues that in 1916 German leaders became more pessimistic about their prospects for military victory in the First World War but responded by *increasing* their war aims. In his view, this risky gamble was an attempt to seek an even bigger victory that would buy off domestic opponents and allow existing elites to stay in power. Other scholars argue that domestic considerations were unimportant, that Germany did not escalate its war aims, that German leaders did not realize in 1916 that they had little chance of winning, or that changes in German leadership accounted for changes in German policies (Goemans 2000). In a third debate, which I recount in the greatest detail below as it is easier to illustrate process tracing with cases with which one is intimately familiar, I have argued that Soviet leaders changed their foreign policies in the late 1980s for ideational reasons. In brief, I maintain that these leaders were reluctant to use force to stem the Central European revolutions of 1989 because they drew the lesson from their failed interventions in Afghanistan and elsewhere that force is not efficacious in achieving long-term political goals like the support of unpopular governments (Bennett 2005). Others have emphasized instead shifts in the balance of power unfavorable to the Soviet Union (Brooks and Wohlforth 2000–1) or changes in the ruling political coalition inside the Soviet Union (Snyder 1987–8).

The adjudication of such competing historical explanations follows a common pattern. We begin with alternative explanations like those in the previous paragraph. We then consider what kinds of evidence, if found, would strengthen or weaken

our confidence in each alternative explanation. Put another way, we consider how strongly each theory anticipates various processes and outcomes, or how surprising it would be from the viewpoint of our prior theoretical expectations to find or to fail to find particular evidence that the hypothesized processes and outcomes took place. Next, we seek out both the expected and the potentially surprising evidence from various sources, taking into account the potential biases that these sources may reflect. After having gathered the available evidence, we update the level of confidence we invest in each of the alternative explanations for the individual case in question. Depending on how this case fits in with our more general theoretical expectations and any earlier empirical findings, the evidence from the case may also lead us to update our confidence in the more general theories lying behind the alternative explanations of the case, and/or our estimate of their scope conditions. This schema for historical explanation is of course oversimplified, and in practice historical explanation can involve many iterations between theoretical conjectures and empirical tests within and across cases. Nonetheless, this logic illustrates much of what is common to both process tracing and Bayesian inference.

## 2 PROCESS TRACING

Process tracing involves looking at evidence within an individual case, or a temporally and spatially bound instance of a specified phenomenon, to derive and/or test alternative explanations of that case. In other words, process tracing seeks a *historical explanation* of an individual case, and this explanation may or may not provide a theoretical explanation relevant to the wider phenomenon of which the case is an instance. Perhaps the researcher will decide that the best explanation for the case is unique or nearly so, such as a finding that a voter supported a candidate who differed from the voter's party, ideology, and issue preferences simply because the candidate was the voter's sister-in-law. Alternatively, the investigator could uncover and test through process tracing a theoretical hypothesis that helps explain a vast range of cases, including even cases that are dissimilar in many respects to the one initially studied. Charles Darwin's theory of evolutionary selection, for example, derived in part from the study of a few species in the Galapagos Islands, is applicable to all living organisms. It is impossible to judge the generalizability of a theory without an understanding of the hypothesized causal mechanisms it entails, and often this understanding is itself developed through the study of an individual case.

Process tracing thus has both inductive (or theory-generating) and deductive (theory-testing) elements. The balance between the two in any particular study depends on the prior state of development of relevant theories on the phenomenon as well as the researcher's state of knowledge about the phenomenon and the case. If

potential alternative explanations of a phenomenon are well established, and if each hypothesis is sufficiently detailed to offer testable predictions on the processes that should have taken place if the hypothesis is to offer a good explanation of the case, then process tracing can proceed largely through deductive theory testing of each alternative explanation (for an example see Schultz 2001, discussed below). If there are no well-specified alternative explanations, however, or if the case is a "deviant" one that does not fit any of the relevant established theories, then process tracing must proceed relatively inductively, through the use of evidence from the case to formulate and possibly later to test new explanations of the case (for an example see Scott 1985). In this regard, while researchers must guard against possible confirmation biases in deriving a theory from a case and then testing it in the same case, it is possible to derive an explanation from a case and then test it against *different and independent evidence* from within that same case. Such evidence provides the possibility of falsifying the new theory as an explanation of the case.

Because process tracing is the technique of looking for the observable implications of hypothesized causal processes within a single case, the researcher engaged in process tracing often looks at a finer level of detail or a lower level of analysis than that of the proposed theoretical explanations.<sup>1</sup> The goal is to document whether the sequence of events or processes within the case fits those predicted by alternative explanations of the case. This is closely analogous to a detective attempting to solve a crime by looking at clues and suspects and attempting to piece together a convincing explanation based on the detailed evidence that bears on means, motives, and opportunity. It is also analogous to a doctor trying to diagnose an illness in an individual patient by taking in the details of the 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). Indeed, all attempts to diagnose or explain individual or historical cases involve process tracing. This is true not only of social events like the Fashoda crisis, the end of the First World War, and the end of the cold war, but of historical events in the hard sciences, like the origins of a particular rock formation (geology), the extinction of the dinosaurs (evolutionary biology), or the origins of the universe (cosmology).

Some critics have suggested that because process tracing frequently involves looking at a finer degree of detail, it can involve a potential "infinite regress" of studying endless "steps between steps between steps" in a case (King, Keohane, and Verba 1994). Indeed, all process-tracing explanations remain provisional, as they could be disproved by further analysis at a finer degree of detail or at a more aggregate level of analysis, but there are pragmatic limits on how much detail a researcher will go into and how continuous an explanation they will seek. These pragmatic limits include the resources available to the researcher, the importance the researcher places on establishing an explanation of the case with a specified degree of confidence, and the rapidity with which the evidence converges in support of a particular explanation.

<sup>1</sup> Process tracing could take place at a higher level of analysis than the hypothesized causal mechanism as well. One could test an economic theory based on hypothesized mechanisms of individual behavior, for example, by examining aggregate behavior in a market.

Moreover, the theoretical importance of a case can change as our research questions evolve—the Fashoda crisis was all but forgotten by everyone except diplomatic historians, for example, until the argument that democracies do not fight one another became an important theoretical conjecture and the Fashoda crisis became a relevant case for testing the alternative mechanisms posited for this hypothesis. Also, a researcher may choose to end or extend their process tracing based on the differential theory-building or policy consequences of a Type I error (rejecting a true explanation) and a Type II error (accepting a false explanation).

In addition, not all pieces of evidence have equal probative value relative to alternative explanations of a case. The more confidently a hypothesis of interest predicts a particular piece of evidence, and the more unlikely that evidence is from the perspective of alternative explanations, the more we increase our confidence in the hypothesis of interest if we do indeed find the evidence it predicted. In this regard, Stephen Van Evera has distinguished among four kinds of tests or evidence representing different combinations of uniqueness and certitude (Van Evera 1997, 31–2). Unique predictions are those accounted for only by one of the theories under consideration, while certain predictions are those that must be unequivocally and inexorably true if an explanation is true. *Hoop tests* involve evidence that is certain but not unique; failure of such a test disqualifies an explanation but passage does not greatly increase confidence in that explanation. Thus, in practice, hoop tests are often used to exclude alternative hypotheses. 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* are unique but not certain; passage strongly affirms an explanation, but failure does not strongly undermine it. As Van Evera notes, a smoking gun in the suspect's hands right after a murder strongly implicates the suspect, but the absence of such a smoking gun does not exonerate a suspect. Smoking gun tests are therefore usually used to instantiate a hypothesis of interest.

*Doubly decisive tests* on evidence that is both unique and certain can provide strong inferences on both the truth and falsity of alternative explanations. Van Evera's example here is that of a bank camera that catches the faces of robbers, thereby implicating the robbers and exonerating all others.

Finally, *straw in the wind tests* are neither unique nor certain and do not provide strong evidence for or against the alternative explanations. Thus, in deciding whether process-tracing evidence is sufficiently convincing to suspend further enquiry into alternative explanations of a case, what matters is not only the importance attached to getting the right explanation, but the kinds of evidence the researcher unearths and the degree of confidence that this evidence allows. A single doubly decisive piece of evidence may suffice, whereas many straw in the wind tests may still leave indeterminacy in choosing between two incompatible explanations of a case. It is not the number of pieces of evidence that matters, but the way in which the evidence stacks up against the alternative explanations.

Process tracing thus works through both affirmation of explanations, through evidence consistent with those explanations, and eliminative induction, or the use

of evidence to cast doubt on alternative explanations that do not fit the evidence in a case. This is why it is important to cast the net widely in considering alternative explanations of a case, and to be meticulous and even-handed in gathering and judging the process tracing evidence on all of the potential explanations of a case that scholars have proposed and that extant theories imply. Other standard injunctions for good process-tracing include the desirability of gathering diverse sources and kinds of evidence and of anticipating and accounting for potential biases in the sources of evidence (George and Bennett 2005; Bennett and Elman 2006). 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 evidence on their behavior or motives.

The example of Heinz Goemans's research on the end of the First World War illustrates these aspects of process-tracing. Goemans uses process-tracing evidence in the manner Van Evera describes to discount four alternative explanations for Germany's policies on the end of the First World War and to instantiate his own explanation. A first alternative explanation, that Germany should have behaved as a unitary actor and responded only to international considerations, fails a hoop test from Goemans's thorough evidence that Germany's war aims increased even as its prospects for victory diminished. A second argument, that Germany was throughout the war irrevocably committed to hegemony, is also undercut by evidence that German war aims in fact increased. Goemans disproves a third argument, that 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 the change in its war aims, begs the question of why Germany replaced its military leaders in the midst of the war (Goemans 2000, 74–5, 93–105).

Goemans then turns to his own explanation, which is that semi-authoritarian governments like Germany's are likely to gamble on war strategies with a small probability of a resounding victory but a high likelihood of abject defeat when they encounter evidence that they are losing a war. Elites in such regimes engage in this kind of behavior, Goemans argues, because for them negotiating an end to the war on modestly concessionary terms is little different from outright loss in the war. In this view, elites fear that domestic opponents could turn widespread unrest over the government's decision to engage in a costly and unsuccessful war into an opportunity to remove elites from power forcibly. Goemans provides smoking gun evidence 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 domestic political reforms would be required to stave off unrest if Germany were to negotiate a concessionary peace. Specifically, the extension of equal voting rights in Prussia "would be worse than a lost war" (Goemans 2000, 114).

### 3 THE LOGIC OF PROCESS TRACING AND BAYESIAN INFERENCE

The logic of process tracing outlined above is strikingly similar to that of Bayesian inference. Like process tracing, Bayesian inference can be used to assess competing explanations of individual historical cases. In contrast, frequentist statistical forms of inference need a sufficiently large sample to get off the ground and face difficult challenges in making inferences about individual cases from aggregate data, a task known as the “ecological inference problem.”<sup>2</sup> Process tracing and Bayesian inference are also closely analogous in their assertions that some pieces of evidence are far more discriminating among competing explanations than others, and they agree on the criteria that determine the probative value of evidence and on the importance of gathering diverse evidence. A third similarity is that both methods proceed through a combination of affirmative evidence on some hypotheses and eliminative induction of other hypothesized explanations that fail to fit the evidence. These similarities point to the potential strengths of the two approaches: It is possible that greatly differing prior expectations about how best to explain a case will converge as evidence accumulates, and depending on the nature of the evidence vis-à-vis the competing explanations, it *may* be possible to make strong inferences even from a few pieces of evidence in a single case. On the other hand, discussions of the limits of Bayesian inference help illuminate the limitations of process tracing as well. These limitations include the potential for underdetermination of choices among competing explanations and ambiguity on how to generalize from a single case even if it is a most- or least-likely case.

It is also important to note several differences between process tracing and Bayesian analysis. Bayesian analysis does not adequately capture the theory-generating side of process tracing, whereby entirely new hypotheses may occur to the researcher on the basis of a single piece of evidence for which no theory had a clear prior. Another difference is that researchers engaged in process tracing typically use Bayesian logic intuitively, rather than attempting to quantify precisely their priors, the specific probabilities associated with hypotheses and evidence, and the updated probabilities of hypotheses in light of evidence. There are also no case-study methods closely analogous to Bayesian statistical analysis of medium to large numbers of cases. Still, case-study researchers can benefit from a clearer understanding of Bayesian analysis and its application to case studies.

The Bayesian approach to theory testing focuses on updating degrees of belief in the truth of alternative explanations.<sup>3</sup> In other words, Bayesians treat statistical

<sup>2</sup> For an optimistic view on addressing and limiting this problem, see King (1997; 1999); for more pessimistic views, see Cho (1998) and Freedman et al. (1998; 1999).

<sup>3</sup> Bayesians often use the metaphor of placing wagers to illustrate how and why beliefs should be updated in light of new information if individuals are to maintain consistent beliefs. In effect, failure to adequately update one’s wagers (beliefs) would allow an arbitrageur to make a series of sure-winner “Dutch Book” bets. Scientists of course do not (usually) make actual monetary wagers based on their

parameters probabilistically, attaching subjective probabilities to the likelihood that hypotheses are true and then updating these probabilities in the light of new evidence. In contrast, frequentist statistics attaches probabilities to the likelihood of getting similar results from repeated sampling of a population. In frequentist statistics, confirming a hypothesis at the “5 percent level” does not mean that there is a 95 percent chance that the hypothesis is true; rather, it means that only one in twenty random samples from the population would lead just by chance to an estimate of the coefficient of the variable in question that is as different from zero as in the sample first tested (Western 1999, 9).

In Bayes’s Theorem, which is derived from Bayes’s work even though he did not state the theorem in exactly this form, the probability that a hypothesis  $H$  is true in light of evidence  $E$  and background knowledge  $K$ , or the conditional probability of  $H$  given  $E$  and  $K$ , is as follows:

$$\Pr(H/E \& K) = \frac{\Pr(H/K) \times \Pr(E/H \& K)}{\Pr(E/K)}$$

In other words, the updated probability of a hypothesis being true in light of new evidence ( $\Pr(H/E \& K)$ ) is equal to the prior probability attached to  $H$  (or  $\Pr(H/K)$ ), times the likelihood of the evidence in light of hypothesis  $H$  and our background knowledge ( $\Pr(E/H \& K)$ ), divided by the prior likelihood of  $E$ . Thus, the more unlikely a piece of evidence is in light of alternatives to explanation  $H$ , the more that evidence increases our confidence that  $H$  is true if the evidence proves consistent with  $H$ .

This corresponds quite closely with the discussion above of how process-tracing evidence from within a case can be used to update our confidence in alternative explanations of the case, and with the argument that not all evidence is of equal probative value. The more surprising process-tracing evidence proves to be for alternative explanations—that is, the more they fail their hoop tests—the more confidence we gain in the remaining hypotheses. The better that evidence fits the explanation of interest—that is, the more smoking gun tests the explanation passes—the more confidence we gain that the theory of interest explains the case. Evidence that sharply raises confidence in some explanations while markedly reducing confidence in others corresponds with Van Evera’s doubly decisive test.

Here again, the analogy to detective work is useful. The sniper team that terrorized the Washington, DC, area in 2002, for example, was finally apprehended after someone called an information hotline and later a local priest claiming to be the sniper. The likelihood of this being true seemed exceedingly low in view of the large number of crank calls on the case, until evidence came to light that substantiated the caller’s claim that the recent shootings were linked to an earlier shooting in Montgomery, Alabama. This evidence in turn generated a number of other leads, each of which was highly unlikely to be true unless the caller was in fact involved in the sniper shootings. When all of these otherwise improbable leads proved to be true, they led in turn to information identifying the car in which the sniper(s) could be found, allowing the

theories, but they do in effect wager their professional reputations and their working hours on the likelihood that hypotheses will prove true.

police to apprehend two suspects who were later convicted of the shootings (Bennett 2006, 341–2).

With enough of the right kind of evidence, even researchers who start out with very different prior probabilities on the alternative explanations should converge toward shared expectations (a phenomenon known as the “washing out of priors”).<sup>4</sup> Similarly, Bayesian and frequentist analyses should converge on similar answers given sufficient evidence, whereas they may diverge when only a few cases or limited evidence are studied. This still leaves the problem of justifying one’s prior probabilities when there is limited evidence, which many view as a key challenge for Bayesian analysis (Earman 1992, 58–9).

Although researchers’ differing priors may or may not converge, the logic of Bayesian inference helps illustrate the common intuition that diversity of evidence is useful in testing hypotheses.<sup>5</sup> Repetitive evidence on the same facet or stage of a process soon loses its probative value, as it quickly becomes predictable or expected and can no longer surprise us or push us to change our priors. Independent evidence on a new or different dimension or stage of a hypothesized process, however, retains the potential for surprising us and forcing us to revise our priors. To borrow from the old adage, if something looks like a duck, walks like a duck, sounds like a duck, and so on, it is probably a duck. If we went on looks alone, however, no matter how many angles of vision we employ we can’t easily tell a duck from a decoy. An example comes from the work of Yuen Foong Khong, who examined the analogies that American leaders used in making and justifying decisions on the Vietnam War. By assessing the analogies these leaders used in private meetings as well as public settings, and finding a close correspondence between the two (with the exception of analogies to Dien Bien Phu, which leaders did not use in public), Khong was able to reject the hypothesis that analogies were used only in public settings for the purposes of instrumental communication to defend choices made in camera for other reasons (Khong 1992, 58–63). Had Khong confined himself to public statements alone, his analysis would have been far less convincing no matter how many such statements he coded, as public statements alone could have been biased by the instrumental purposes of the actors making these statements. As in process tracing, it is not the

<sup>4</sup> There are methods in Bayesian statistical analysis for assessing the sensitivity of results to different priors, and Bayesian Markov Chain Monte Carlo analysis can use diffuse priors so that the results are not strongly affected by the priors (Western 1999, 11; Jackman 2004, 484, 488). There is no equivalent in the informally Bayesian techniques of process tracing, although the convergence on an explanation of a case by researchers who started out with different priors could be construed as an example of the washing out of priors.

<sup>5</sup> For arguments that Bayesianism helps explain the value of diversity of evidence, see Earman (1992, 78–9), Hellman (1997, 202–9); for a critique, see Kruse (1999, 171–4). Even if successful, Bayesian arguments on this point create a potential paradox, known as the problem of “old evidence.” The Bayesian theorem suggests that evidence already known cannot have additional probative value, as it has already been incorporated into the prior estimate of the likely truth of a hypothesis. This means that evidence can have probative value for one person who was previously ignorant of it, but not for another, who knew it already. This can be viewed as either a problem for Bayesianism, or as consistent with the intuition that priors are to some degree subjective and that researchers must guard against confirmation biases that attend old evidence. See Earman (1992, 119–20, 132).

number of cases or of pieces of evidence that matters, but the discriminating power and diversity of evidence vis-à-vis the alternative hypotheses under consideration.

This is quite different from the frequentist logic of inference, and it helps explain why the “degrees of freedom” critique sometimes leveled at case-study methods is woefully misguided. In frequentist statistical inference, the term “degrees of freedom” refers to the number of observations or cases minus the number of estimated parameters, or characteristics of the population being studied. In frequentist studies, this number is crucial, as the higher the degrees of freedom, the more powerful the test is at determining whether the observed variance could have been brought about by chance, and statistical studies simply cannot be done when there are negative degrees of freedom. This has at times led researchers trained in the frequentist tradition to worry that case studies that involve more parameters than cases are inherently indeterminate research designs. Any particular case-study design may in fact prove to be indeterminate in a deeper sense—indeed, as discussed below, Bayesian logic suggests that all research designs are on some level indeterminate—but the determinacy of any particular case-study design is not closely associated with the number of cases or parameters being studied.

The noted methodologist Donald Campbell got this intuition at least partly right when he set out to “correct some of my own prior excesses in describing the case study approach” and reconsidered his earlier critique of case-study methods for allegedly lacking degrees of freedom. Campbell argued that the theory a researcher uses to explain, for example, cultural differences “also generates predictions or expectations on dozens of other aspects of the culture, and he [*sic*] does not retain the theory unless most of these are also confirmed. In some sense, he [*sic*] has tested the theory with degrees of freedom coming from the multiple implications of any one theory” (Campbell 1975). Notice, however, that Campbell is still thinking in relatively frequentist terms about the nature of evidence in focusing on the *numbers* of expectations that are or are not met rather than the *probative value and diversity of evidence* relative to the competing hypotheses, as Bayesian logic suggests. The probative value of evidence does not depend on the number of pieces of evidence or the number of cases, but on how the evidence stacks up against competing explanations. A single piece of the right kind of evidence can establish one explanation of a case as highly likely to be true and all other proposed explanations as likely to be wrong, whereas thousands of pieces of evidence that do not discriminate between competing explanations may have no probative value at all. One piece of doubly decisive evidence can outweigh many less decisive bits of seemingly contradictory evidence.

Another parallel is that, in contrast to approaches that emphasize either confirmatory or disconfirming evidence, both process tracing and Bayesian inference proceed through a combination of affirmation of hypotheses consistent with the evidence and eliminative induction of hypothesis that do not fit. As is evident in Bayes’s theorem, evidence that is highly unlikely in terms of the alternative hypotheses can raise our posterior estimate of the likely truth of a hypothesis just as evidence consistent with that hypothesis can. This is the root of Sir Arthur Conan Doyle’s advice, through his fictional detective Sherlock Holmes, that an investigation “starts upon

the supposition that when you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth. It may well be that several explanations remain, in which case one tries test after test until one or other of them has a convincing amount of support" (Doyle 1927, 1011, in "The Blanched Soldier" in *The Case-Book of Sherlock Holmes*; see also Earman 1992, 163–5 on eliminative induction).

Kenneth Schultz provides an excellent example of both affirmative evidence and eliminative induction in his analysis of the 1898 Fashoda crisis between Britain and France. The Fashoda crisis arose over the convergence of British and French expeditionary forces on the town of Fashoda as they raced to lay claim to the Upper Nile Valley. Schultz lays out three alternative explanations that scholars have offered for why the crisis was resolved without a war when France chose to back down from its claims. Neorealists have argued that France backed down short of war simply because Britain's military forces were far stronger, both in the region and globally. 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 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. The British public and British leaders, far from being conciliatory, as traditional democratic peace theories based on the restraining power of democratic norms and institutions would suggest, were belligerent in their rhetoric and actions toward France throughout the crisis (Schultz 2001, 180–3).

Schultz then turns to his own explanation, which is that democratic institutions force democratic leaders to reveal private information about their intentions, making it difficult for democratic leaders to bluff in some circumstances but also making their 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 in its favor. 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 made public commitments to an intransigent position, but he was quickly undermined by public evidence of apathy and even outright opposition by other domestic political actors (Schultz 2001, 193). After this costly signaling by both sides revealed that Britain had a greater willingness as well as capability to fight for the Upper Nile, within a matter of days France began to back down, leading to a resolution of the crisis in Britain's favor. The close timing of these events, following in the sequence predicted by Schultz's theory, provides strong evidence for his explanation.

Discussions of Bayesian inference also illuminate the limitations of both Bayesianism and process tracing. Foremost among these is the underdetermination thesis (sometimes referred to as the Quine–Duhem problem after the two philosophers of science who called attention to it, Pierre Duhem and Willard Van Orman Quine).

In this view, the choice among competing theories is always underdetermined by evidence, as it is not entirely clear whether evidence falsifies a theory itself or one of the theory's (often implicit) "auxiliary hypotheses" about contextual variables, methods of observation and measurement, and so on. It is also not clear which of the potentially infinite alternative explanations deserves serious consideration, and by definition we cannot consider potentially true alternative hypotheses that we simply have not yet conceived. In Bayes's theorem, these problems arise in the term  $K$ , which represents the likelihood of evidence in light of our background knowledge of the alternative hypotheses (or, put another way, the probability of observing  $E$  if  $H$  is not true). This term is sometimes called the "catchall factor" because it is in effect a grab-bag of all hypotheses other than  $H$ , including even those not yet proposed. As it is not actually possible to consider all the potential alternative explanations, Bayesians have gravitated to the pragmatic advice, similar to that offered by proponents of process tracing, that one should cast the net for alternative explanations widely, considering even those explanations that have few proponents if they make different predictions from those of the hypothesis of interest (Earman 1992, 84–5; for a critique see Kruse 1999, 165–71). This remains at best a provisional approach, however, which is one reason that Bayesians never attach 100 percent probability to the likely truth of theories.

Another version of the underdetermination problem is that the evidence might not even allow us to choose decisively between two extant hypotheses that give incompatible explanations. Differences between priors, in other words, may not wash out if there is not adequate evidence, bringing us back to the problem that priors themselves lack any clear justification. As argued above, the severity of this problem in any particular study depends on the nature of the competing hypotheses vis-à-vis the evidence, not on the number of cases or variables—indeed, underdetermination of this sort can be a problem even with many cases and few variables. With the right kind of evidence, some studies may achieve a high degree of confidence in a particular explanation, but there is no general assurance that any given case study will achieve this result.

A third challenge for both Bayesianism and process tracing is that of generalizing from one or a few cases. Bayesian logic can help to some degree with this problem. For example, case-study methodologists have long argued, consistent with Bayesianism, that if a hypothesis appears to accurately explain a tough test case which, a priori, it looked "least likely" to explain, then the hypothesis is strongly affirmed. Conversely, failure to explain a "most likely" case strongly undermines our confidence in a hypothesis. It remains unclear, however, whether these inferences should apply only to the case being studied, to very similar cases to the one studied, or to a broader range of more diverse cases. This kind of inference depends on both our knowledge of the hypothesis, which may itself provide clues on its scope conditions, and our knowledge of the population of cases for which the hypothesis is relevant. These kinds of information must come from either prior knowledge or further study (whether Bayesian or frequentist) of more cases.

To take an example from actual research, my colleagues and I concluded from our study of burden sharing in the 1991 Persian Gulf War that one hypothesis had proved

true even in the least-likely cases of Germany and Japan. This was the hypothesis that countries would contribute to an alliance if they were dependent on the alliance leader for their security. In the cases of Germany and Japan, this proved true even though every other hypothesis—the collective action/free riding hypothesis, the domestic politics hypothesis (political opposition at home and a weak state on foreign policy), the “balance of threat” hypothesis—pointed against a substantial contribution by either country. This finding suggested that the alliance dependence hypothesis generally deserved more weight than it had been accorded in the alliance literature, which up to that time was dominated by the collective action hypothesis. In the absence of wider statistical or qualitative studies of more episodes of alliance behavior, however, we could not conclude precisely on the conditions under which or the extent to which an increase in alliance dependence would lead to a specific increase in burden-sharing commitments (Bennett, Lepage, and Unger 1997).

Finally, there are areas in which Bayesian inference and process tracing are simply different. Bayesianism does not have any equivalent for the generation of new hypotheses from the close process-tracing study of a case. As John Earman has noted, “data-instigated hypotheses call for literally ad hoc responses. When new evidence suggests new theories, a non-Bayesian shift in the belief function may take place” (Earman 1992, 100; see also Hellman 1997, 215). Another issue to which Bayesians have paid less attention than qualitative methodologists is the problem of potential biases in information sources. It is certainly consistent with Bayesianism to discount the likely truth value of evidence that comes from biased or incomplete sources, but most discussions of Bayesian inference do not focus on this problem, perhaps because they usually address examples in the natural sciences in which potential measurement error, rather than intentional obfuscation of evidence by goal-oriented actors, is the most common problem with evidence.

#### 4 AN EXTENDED EXAMPLE: DEBATES ON THE END OF THE COLD WAR

There are many excellent examples of process tracing and Bayesian inference in political science. In the international relations subfield alone these would include but certainly not be limited to books by Drezner (1999), Eden (2006), Evangelista (2002), George and Smoke (1974), Homer-Dixon (2001), Khong (1992), Knopf (1998), Larson (1989), Moravcsik (1998), Sagan (1993), Shafer (1988), Snyder (1984; 1993), Walt (1996), and Weber (1991).<sup>6</sup> I focus here on the historiographical debate on the peaceful

<sup>6</sup> For brief descriptions of the research designs by Drezner, George and Smoke, Homer-Dixon, Khong, Knopf, Larson, Owen, Sagan, Shafer, Snyder, and Weber, see George and Bennett (2005, 118–19, 194–7, 302–25); on Evangelista, see Bennett and Elman (2007).

end of the cold war, and particularly the absence of Soviet military interventions in the Eastern European revolutions of 1989. I have contributed to this debate, and it is one on which scholars' prior theoretical expectations have only partially converged. I introduce it here not to convince readers of my own side of this debate, which would require a far more detailed analysis of the competing views and evidence, but to illustrate the kind of Bayesian inference from process-tracing evidence that is involved in such judgments. I also use this example because it is far easier to reconstruct the process-tracing involved in one's own research than to dissect process tracing done by other researchers. I focus here on three of the most prominent explanations for the nonuse of force in 1989: the realist explanation, which emphasizes the changing material balance of power; the domestic politics explanation, which focuses on the changing nature of the Soviet Union's ruling coalition; and the ideational explanation, which highlights the lessons Soviet leaders drew from their recent experiences in using force.

Stephen Brooks and William Wohlforth (Brooks and Wohlforth 2000–1; see also Wohlforth 1994–5; Oye 1996) have constructed the most comprehensive realist/balance of power explanation for Soviet restraint in 1989. Brooks and Wohlforth argue that the decline in Soviet economic growth rates in the 1980s, combined with the Soviet Union's “imperial overstretch” in Afghanistan and its high defense spending burden, was the driving force in Soviet foreign policy retrenchment in the late 1980s. In particular, these authors argue, 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.

Jack Snyder has been the foremost proponent of a second hypothesis focused on Soviet domestic politics (Snyder 1987–8). Snyder argues that the long-term shift in the Soviet economy from extensive development (focused on basic industrial goods) to intensive development (involving more sophisticated and information-intensive goods and services) brought about a shift in the ruling Soviet coalition from the military-heavy industry-party complex to light industry and the intelligentsia. This led the Soviet Union to favor improved ties to the West to gain access to technology and trade, which would have been undermined had the Soviet Union used force in 1989. Snyder has not directly applied this argument to Soviet restraint in the use of force in 1989, though he did continue to view sectoral material interests as the driving factor in Soviet policy after 1989 (Snyder 1990).

I have emphasized a third “learning” explanation, namely that Soviet leaders drew lessons from their recent unsuccessful military interventions in Afghanistan and elsewhere that made them doubt the efficacy of using force to resolve long-term political problems (Bennett 1999; 2003; 2005; see also English 2000; 2002; Checkel 1997; Stein 1994).

While scholars agree that the variables highlighted by all three of these hypotheses contributed to the nonuse of force in 1989, there remains considerable disagreement



on their relative causal weight, the ways in which they interacted, and the counterfactuals that each hypothesis could plausibly sustain. 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" (Brooks and Wohlforth 2002, 94). In contrast, I find this "standard view" convincing and argue that although the Soviet Union's material decline relative to the West created pressures for foreign policy change, the timing and direction of the foreign policy changes that resulted were greatly influenced by the lessons Soviet leaders drew from their recent failures in the use of force in Afghanistan and elsewhere.

How are we to judge the competing claims of each hypothesis? Nina Tannenwald has offered three relevant process-tracing tests (Tannenwald 2005): First, 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? Second, did material change precede or follow ideational change? Third, do material or ideational factors better explain which ideas won out? A brief overview of the evidence on each of these points reveals "smoking gun" evidence for both the material and ideational explanations, and a failed "hoop test" for one variant of the material explanation; on the whole, it appears that evidence for sectoral material interests is weakest.

On Tannenwald's first test, the correlation of policy positions with material versus ideational variables, there is some evidence for each explanation, mostly of the "straw in the wind" variety that can only be briefly summarized here. Brooks and Wohlforth argue, citing Soviet Defense Minister Yazov and others, that Soviet conservatives and military leaders largely 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 (Brooks and Wohlforth 2000-1). Robert English, pointing to other statements by Soviet conservatives indicating opposition to Gorbachev's foreign policies, concludes instead 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). It is hard to conceive of what "smoking gun" or "hoop test" evidence would look like on this issue, however, as the policy views of any one individual, even an individual as historically important as Gorbachev, cannot definitively show that material incentives rather than ideas were more important in driving and shaping changes in Soviet policies.

More definitive is the hoop-type evidence against Snyder's sectoral interest group explanation: Although Soviet military leaders did at times argue 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, these actors did *not* argue, even after they had already fallen from power and had little to lose, that force should have been used to prevent the dissolution of the Warsaw Pact in 1989 (Bennett 2005, 104). Indeed, military leaders were among the early skeptics on the

use of force in Afghanistan, and many prominent officers with personal experience in Afghanistan resigned their commissions rather than participate in the 1994-7 Russian intervention in Chechnya (Bennett 1999, 339-40).

Tannenwald's second test concerns the timing of material and ideational change. Brooks and Wohlforth have not stated precisely the timeframe within which they believe material decline would have allowed or compelled Soviet foreign policy change, stating only that material incentives shape actions over the "longer run" (2002, 97). The logic of their argument, however, suggests that the Soviet Union could have profitably let go of its Eastern European empire in 1973, by which time nuclear parity guaranteed the Soviet Union's security from external attack and high energy prices meant that the Soviet Union could have earned a higher price for its oil and natural gas on world markets than it got 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 the ideational explanation coincides much more closely with changes in Soviet foreign policy. Soviet leaders were quite optimistic about the use of force in the developing world in the late 1970s, despite slow Soviet economic growth, but they became far more pessimistic regarding the efficacy of force as their failure in Afghanistan became more apparent (Bennett 1999). Moreover, changes in Soviet leaders' public statements generally preceded changes in Soviet foreign policy, suggesting that ideational change, rather than material interests justified by ad hoc and post hoc changes in stated ideas, were the driving factor (Bennett 1999, 351-2).

The most definitive process-tracing evidence, however, concerns the third question of why some ideas won out over others. Here, there is strong evidence that both material and ideational factors played a role, but one variant of the material explanation appears to have failed 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 the latter was predominant in the mind of the key actor, Mikhail Gorbachev. 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 the Soviet economy's 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.

Finally, counterfactual analysis can shed light on analysts' Bayesian expectations and the need to update them. Jack Snyder carefully and conscientiously outlined in early 1988 the (then) counterfactual future events that would in his view have led to a resurgence of the Stalinist coalition of the military and heavy industry. If Gorbachev's reforms were discredited through 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 argued, the rise of anti-reform Soviet leaders was much more likely. As it turned out, every one of these conditions except for the deployment of a working SDI system was over-fulfilled within two years, yet apart from the thoroughly unsuccessful coup attempt of 1990, Soviet hardliners never came close to regaining power.

It is also worth considering the counterfactual implications of the material and ideational arguments. The counterfactual of the Brooks–Wohlforth argument is that had Gorbachev's reforms succeeded in dramatically improving Soviet economic production by 1989, the Soviet Union would have been more likely to use force against the revolutions in Eastern Europe. Brooks and Wohlforth have never gone so far as to assert or defend this claim, however. The counterfactual implication of my own argument, and one that I find eminently plausible, is that if Soviet displays and uses of force in Poland in 1980–1, in Afghanistan in 1979 and after, and in other conflicts in the 1970s and 1980s had been considered successes by Soviet leaders in 1989, they would have been far more likely to have threatened and used force to stop the revolutions of 1989. The (apparently) different degrees of belief that we attach to the counterfactual implications of our arguments of course do not constitute evidence on the truth of these arguments, but they do reinforce the point that Bayesian logic's focus on updating degrees of belief through evidence is an apt description of the logic of process tracing as well.

## 5 CONCLUSION

The logic of Bayesian inference illuminates both the strengths and limitations of process tracing. Among the strengths of both approaches is that it is possible to make strong inferences in just one or a few cases, based on one or a few pieces of the right kind of evidence, if this evidence strongly discriminates between alternative hypotheses in the ways discussed above. Thus, the "Degrees of Freedom" problem

is inapplicable to process tracing, even though the more fundamental problem of underdetermination remains. Both approaches are congruent as well, in that they proceed by both affirmation and eliminative induction, that they view some kinds of evidence as far more probative than other kinds of evidence, and that they stress the importance of obtaining diverse evidence on the phenomenon under study.

As for weaknesses, both Bayesianism and process tracing face the problem that prior theoretical expectations may have no absolute justification. With the right kind of evidence differing priors may converge, but in the absence of such evidence they will not. Bayesian logic also serves as a useful reminder that the explanations derived from process tracing are always provisional, which is another way of framing the problem of underdetermination. Because we can never have a full accounting of the background factor of all possible alternative explanations, we should never let our confidence in the likely truth of an explanation equal 100 percent. Moreover, even if the evidence gives us a high degree of confidence in an explanation of a case, Bayesian logic offers only partial help on the challenge of generalizing from a case. Theories that succeed in their least-likely cases may deserve greater weight and scope, and those that fail in most-likely cases may deserve to be rejected or narrowed, but beyond this general conclusion it remains unclear whether or how Bayesian logic can assist in identifying the populations to which an explanation might be applicable, especially in the absence of prior knowledge about such potentially relevant populations.

Bayesianism and process tracing do not overlap completely. There is no Bayesian equivalent for generating a completely new hypothesis from the close process-tracing study of a case. As the example of Darwin's inferences on evolution reminds us, there may also be ways of generalizing from a new hypothesis derived from a case, and from prior knowledge about populations that share some features of the case, that prove to have little to do with Bayesianism. Yet for all their differences, perhaps including those not yet discussed nor discovered (the catchall factor!), the similarities between Bayesian and process-tracing analyses of individual cases, from medical diagnoses to studies of war and peace, reveal the uses and limits of both.

## REFERENCES

- BENNETT, A. 1999. *Condemned to Repetition? The Rise, Fall, and Reprise of Soviet-Russian Military Interventionism 1973–1996*. Cambridge, Mass.: MIT Press.
- 2003. Trust bursting out all over: the Soviet side of German unification. Pp. 175–204 in *Cold War Endgame*, ed. W. Wohlforth. University Park: Pennsylvania State University Press.
- 2005. The guns that didn't smoke: ideas and the Soviet non-use of force in 1989. *Journal of Cold War Studies*, 7: 81–109.
- 2006. Stirring the frequentist pot with a dash of Bayes. *Political Analysis*, 14: 339–44.
- and ELMAN, C. 2006. Qualitative research: recent developments in case study methods. *Annual Review of Political Science*, 9: 455–76.
- 2007. Case study methods in the international relations subfield. *Comparative Political Studies*, 40: 170–95.

- BENNETT, A., LEFGOLD, J., and UNGER, D. (eds.) 1997. *Friends in Need: Burden Sharing in the Persian Gulf War*. New York: St Martin's Press.
- BRADY, H., and COLLIER, D. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Savage, Md.: Rowman and Littlefield.
- BROOKS, S., and WOHLFORTH, W. 2000-1. Power, globalization, and the end of the Cold War: reevaluating a landmark case for ideas. *International Security*, 25: 5-53.
- 2002. From old thinking to new thinking in qualitative research. *International Security*, 26: 93-111.
- BROWN, A. 1996. *The Gorbachev Factor*. Oxford: Oxford University Press.
- CAMPBELL, D. 1975. Degrees of freedom and the case study. *Comparative Political Studies*, 8: 178-85.
- CHECKEL, J. 1997. *Ideas and International Political Change: Soviet/Russian Behavior and the End of the Cold War*. New Haven, Conn.: Yale University Press.
- CHO, W. T. 1998. If the assumption fits...: a comment on the King ecological inference solution. *Political Analysis*, 7: 43-163.
- DOYLE, A. C. 1927. *The Casebook of Sherlock Holmes*. London: John Murray.
- DREZNER, D. 1999. *The Sanctions Paradox: Economic Statecraft and International Relations*. Cambridge: Cambridge University Press.
- EARMAN, J. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press.
- EDEN, L. 2006. *Whole World on Fire: Organizations, Knowledge, and Nuclear Weapons Devastation*. Ithaca, NY: Cornell University Press.
- ENGLISH, R. 2000. *Russia and the Idea of the West: Gorbachev, Intellectuals, and the End of the Cold War*. New York: Columbia University Press.
- 2002. Power, ideas, and new evidence on the Cold War's end: a reply to Brooks and Wohlforth. *International Security*, 26: 70-92.
- EVANGELISTA, M. 2002. *Unarmed Forces: The Transnational Movement to End the Cold War*. Ithaca, NY: Cornell University Press.
- FREEDMAN, D. A., KLEIN, S. P., OSTLAND, M., and ROBERTS, M. R. 1998. Review of *A Solution to the Ecological Inference Problem* by G. King. *Journal of the American Statistical Association*, 93: 1518-22.
- OSTLAND, M., ROBERTS, M. R., and KLEIN, S. P. 1999. Response to King's comment. *Journal of the American Statistical Association*, 94: 355-7.
- GEORGE, A. L., and BENNETT, A. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass.: MIT Press.
- and SMOKE, R. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. New York: Columbia University Press.
- GILL, C. J., SABIN, L., and SCHMID, C. H. 2005. Clinicians as natural Bayesians. *British Medical Journal*, 330: 1080-3.
- GOEMANS, H. 2000. *War and Punishment: The Causes of War Termination and the First World War*. Princeton, NJ: Princeton University Press.
- HELLMAN, G. 1997. Bayes and beyond. *Philosophy of Science*, 64: 191-221.
- HOMER-DIXON, T. 2001. *Environment, Scarcity, and Violence*. Princeton, NJ: Princeton University Press.
- JACKMAN, S. 2004. Bayesian analysis for political research. *Annual Reviews of Political Science*, 7: 483-505.
- KHONG, Y. F. 1992. *Analogies at War: Korea, Munich, Dien Bien Phu and the Vietnam Decisions of 1965*. Princeton, NJ: Princeton University Press.
- KING, G. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

- 1999. The future of ecological inference research: a comment on Friedman et al. *Journal of the American Statistical Association*, 94: 352-5.
- KEOHANE, R., and VERBA, S. 1994. *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
- KNOPE, J. 1998. *Domestic Society and International Cooperation: The Impact of Protest on U.S. Arms Control Policy*. Cambridge: Cambridge University Press.
- KRUSE, M. 1999. Beyond Bayesianism: comments on Hellman's "Bayes and beyond." *Philosophy of Science*, 66: 165-74.
- LARSON, D. W. 1989. *Origins of Containment*. Princeton, NJ: Princeton University Press.
- MCKEOWN, T. 2003. Case studies and the statistical world view. *International Organization*, 53: 161-90.
- MORAVCSIK, A. 1998. *The Choice for Europe: Social Purpose and State Power from Messing to Maastricht*. Ithaca, NY: Cornell University Press.
- OYE, K. 1996. Explaining the end of the Cold War: morphological and behavior adaptations to the nuclear peace? Pp. 57-83 in *International Relations Theory and the End of the Cold War*, ed. R. N. Lebow and T. Risse-Kappen. New York: Columbia University Press.
- SAGAN, S. 1993. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press.
- SCHULTZ, K. A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.
- SCOTT, J. C. 1985. *Weapons of the Weak*. New Haven, Conn.: Yale University Press.
- SHAFFER, D. M. 1988. *Deadly Paradigms: The Failure of U.S. Counterinsurgency Policy*. Princeton, NJ: Princeton University Press.
- SNYDER, J. 1984. *The Ideology of the Offensive: Military Decision Making and the Disasters of 1914*. Ithaca, NY: Cornell University Press.
- 1987-8. The Gorbachev revolution: a waning of Soviet expansionism? *International Security*, 12: 93-131.
- 1990. Averting anarchy in the new Europe. *International Security*, 14: 5-41.
- 1993. *Myths of Empire: Domestic Politics and International Ambition*. Ithaca, NY: Cornell University Press.
- STEIN, J. G. 1994. Political learning by doing: Gorbachev as uncommitted thinker and motivated learner. *International Organization*, 48: 155-83.
- TANNENWALD, N. 2005. Ideas and explanation: advancing the theoretical agenda. *Journal of Cold War Studies*, 7: 13-42.
- VAN EVERA, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- WALT, S. 1996. *Revolution and War*. Ithaca, NY: Cornell University Press.
- WEBER, S. 1991. *Cooperation and Discord in U.S.-Soviet Arms Control*. Princeton, NJ: Princeton University Press.
- WESTERN, B. 1999. Bayesian analysis for sociologists: an introduction. *Sociological Methods and Research*, 28: 7-34.
- WOHLFORTH, W. 1994-5. Realism and the end of the Cold War. *International Security*, 19: 91-129.
- ZELIKOW, P., with RICE, C. 1995. *Germany Unified and Europe Transformed: A Study in Statecraft*. Cambridge, Mass.: Harvard University Press.

## CHAPTER 31

CASE-ORIENTED  
CONFIGURATIONAL RESEARCH:  
QUALITATIVE  
COMPARATIVE  
ANALYSIS (QCA),  
FUZZY SETS, AND  
RELATED  
TECHNIQUES

BENOÎT RIHOUX

With the support of the Fonds National de la Recherche Scientifique—FNRS (Belgium)—FRFC fund. Many ideas in this chapter have been elaborated together with Charles Ragin, and discussed at length with other contributors of an edited textbook on this topic (Rihoux and Ragin 2008).

Most of the points in this chapter, as well as practical descriptions and discussions of the techniques, are further elaborated in two textbooks (Rihoux and Ragin 2008; Schneider and Wagemann 2007; forthcoming). Some software instructions are also available through <http://www.compass.org>—“software” links).

## 1 INTRODUCTION

QUALITATIVE comparative analysis (or QCA<sup>1</sup>; Ragin 1987) and linked techniques such as fuzzy sets (Ragin 2000; and this volume) were developed for the analysis of small- and intermediate-*N* data-sets, typical of those used by researchers in comparative politics and related disciplines. These techniques are designed to unravel causal complexity by applying set-theoretic methods to cross-case evidence. Their central goal is to mimic some of the basic analytic procedures that comparative researchers use routinely when making sense of their cases. The key difference between QCA and traditional case-oriented methods is that with QCA it is possible to extend these basic analytic procedures to the examination of more than a handful of cases. There is no procedural limit on the number of cases that can be studied using QCA.

This chapter offers a conceptually oriented introduction to QCA, also discussing some key technical issues. First, two analytic procedures commonly used by comparative researchers are laid out and contrasted with correlational analysis, the main analytical engine of mainstream quantitative social science. Second, a short description of the state-of-the-art of QCA applications is provided, in terms of discipline, types of cases, models, combinations with other methods, and software development. Next, different uses of QCA are outlined, as well as generic “best practices.” Fourth, some key recent evolutions are presented: on the one hand the development, beyond dichotomous “crisp set” QCA (csQCA), of multi-value QCA (mvQCA), fuzzy sets, and fuzzy-set QCA (fsQCA), and on the other hand technical advances and refinements in the use of the techniques. Finally, concluding reflections are offered as to expected developments, upcoming innovations, remaining challenges, expansion of fields of application, and cross-fertilization with other approaches.

2 THE DISTINCTIVENESS OF  
COMPARATIVE RESEARCH

Researchers in comparative politics and related fields often seek to identify commonalities across cases, focusing on a relatively small number of purposefully selected cases. There are two analytic strategies central to this type of research. The first one is to examine cases sharing a given *outcome* (e.g. consolidated third-wave democracies)

<sup>1</sup> Hereafter, QCA is used generically to refer to both csQCA (crisp-set QCA, dichotomous), mvQCA (multi-value QCA), and fsQCA (fuzzy-set QCA, a variant linking fuzzy sets to QCA procedures). All are grouped, together with fuzzy sets, under the “configurational comparative methods” label (Rihoux and Ragin 2008).

and to attempt to identify their shared *conditions*<sup>2</sup> (e.g. the possibility that they share presidential systems). The second one is to examine cases sharing a specific condition or, more commonly, a specific combination of conditions, and to assess whether or not these cases exhibit the same outcome (e.g. do cases that combine party fractionalization, a weak executive, and a low level of economic development all suffer democratic breakdown?). Both strategies are set theoretic in nature. The first is an examination of whether instances of a specific outcome constitute a subset of instances of one or more causal conditions. The second is an examination of whether instances of a specific causal condition or combination of causal conditions constitute a subset of instances of an outcome.

Both strategies are methods for establishing *specific* connections. If it is found, for example, that all (or nearly all<sup>3</sup>) consolidated third-wave democracies have presidential systems, then a specific connection has been established between presidentialism and consolidation—assuming this connection dovetails with existing theoretical and substantive knowledge. Likewise, if it is found that all (or nearly all) third-wave democracies that share a low level of economic development, party fractionalization, and a weak executive failed as democracies, then a specific connection has been established between this combination of conditions and democratic breakdown. Establishing specific connections is not the same as establishing correlations. For example, assume that the survival rate for third-wave democracies with presidential systems is 60 percent, while the survival rate for third-wave democracies with parliamentary systems is 35 percent. Clearly, there is a correlation between these two aspects conceived as variables (presidential versus parliamentary system and survival versus failure). However, the evidence does not come close to approximating a set-theoretic relation: There is evidence of correlation (i.e. a general connection), but not of a specific connection between presidential systems and democratic survival in this example.

As explained in Ragin (2000), the first analytic strategy—identifying conditions shared by cases with the same outcome—is appropriate for the assessment of necessary conditions. The second—examining cases with the same causal conditions to see if they also share the same outcome—is suitable for the assessment of sufficient conditions, especially sufficient combinations of conditions. Establishing conditions that are necessary or sufficient is a long-standing interest of comparative researchers (see, e.g., Goertz and Starr 2002). However, the use of set-theoretic methods to establish explicit connections does not necessarily entail the use of the concepts or the language of necessity and sufficiency, or any other language of causation. A researcher might observe, for example, that instances of democratic breakdown are all ex-colonies without drawing any causal connection from this observation. Demonstrating explicit connections is important to social scientists, whether or not

<sup>2</sup> The term *condition* is used generically to designate an aspect of a case that is relevant in some way to the researcher's account or explanation of some *outcome* exhibited by the case. Note that it is not an "independent variable" in the statistical sense.

<sup>3</sup> Neither strategy expects or depends on perfect set-theoretic relations (see Ragin, this volume).

Table 31.1. Cross-tabulation of presence/absence of an outcome against presence/absence of a condition

	Condition absent	Condition present
Outcome present	Cell 1: cases here undermine researcher's argument	Cell 2: cases here support researcher's argument
Outcome absent	Cell 3: cases here support researcher's argument	Cell 4: cases here undermine researcher's argument

they are interested in demonstrating causation. In fact, qualitative analysis in the social sciences is centrally concerned with establishing specific connections.

Correlational methods are not well suited for studying specific connections (Ragin 2000). This mismatch is clearly visible in the simplest form of variable-oriented analysis, the  $2 \times 2$  cross-tabulation of the presence/absence of an outcome against the presence/absence of an hypothesized cause (Table 31.1).

The correlation (used in the generic sense) focuses simultaneously and equivalently on the degree to which instances of the cause produce instances of the outcome (the number of cases in cell 2 relative to the sum of cells 2 and 4) *and on* the degree to which instances of the absence of the cause are linked to the absence of the outcome (the number of cases in cell 3 relative to the sum of cells 1 and 3). In short, it is an omnibus statistic that rewards researchers for producing an abundance of cases in cell 2 and/or cell 3 and penalizes them for depositing many cases in cell 1 and/or cell 4. Thus, it is a good tool for studying general (correlational) connections.

A researcher interested in specific connections, however, focuses only on some specific components of the information that is conflated in a correlation. For example, comparative researchers interested in causally relevant conditions shared by instances of an outcome would focus on cells 1 and 2. Their goal would be to identify conditions that deposit as few cases as possible in cell 1. Likewise, researchers interested in whether cases that are similar with respect to conditions experience the same outcome would focus on cells 2 and 4. Their goal would be to identify combinations of causal conditions that deposit as few cases as possible in cell 4. It is clear from these examples that the correlation has two major shortcomings when viewed from the perspective of specific connections: (1) it attends only to relative differences (e.g. relative survival rates of presidential versus parliamentary systems); and (2) it conflates different kinds of set-theoretic assessment.

As the bivariate correlation is the foundation of most forms of conventional quantitative social research, including some of the most sophisticated forms of variable-oriented analysis practiced today (Ragin and Sonnett 2004), these sophisticated quantitative techniques eschew the study of explicit connections. QCA, by contrast, is centrally concerned with specific connections. It is grounded in formal logic (Boolean algebra and set-theoretic language), and thus is ideally suited for identifying key set-theoretic relations. An especially useful feature of QCA is its capacity for analyzing

complex causation (“multiple conjunctural causation”), defined as a situation where a given outcome may follow from several different combinations of conditions—different causal “paths.” Let us consider, for instance, a qualitative outcome such as being fired (or not) from a job. There are different things one could do to get fired: stealing, showing up late, defaming, etc., or all these things combined. The point is that each one of these actions leads one to be fired, and they are all different ways to be fired. In this instance, a single model of the process would be a gross misrepresentation of what happens empirically. It wouldn’t mean much to estimate the relative influence of those different actions on the outcome of “being fired,” because people get fired for different reasons (or different combinations of reasons), and not because they scored little on stealing plus a little on showing up late at work, etc.

To sum up: With QCA, by examining the fate of cases with different combinations of causally relevant conditions, it is possible to identify the decisive recipes and thereby unravel causal complexity. Note, in this respect, that QCA does not take on board some core assumptions of the mainstream quantitative approach: It does not assume linearity, nor additivity, and it assumes equifinality (i.e. different paths can lead to the same outcome) instead of unit homogeneity (a given factor is assumed to have the same effect on the outcome across all cases) (Rihoux and Ragin 2008; Schneider and Wagemann forthcoming).

The key analytic tool for analyzing causal complexity using QCA is the truth table. It lists the logically possible combinations of conditions and the outcome associated

Table 31.2. Truth table with four conditions (A, B, C, and D) and one outcome (Y)

A	B	C	D	Y <sup>a</sup>
no	no	no	no	no
no	no	no	yes	?
no	no	yes	no	?
no	no	yes	yes	?
no	yes	no	no	no
no	yes	no	yes	no
no	yes	yes	no	?
no	yes	yes	yes	no
yes	no	no	no	?
yes	no	no	yes	?
yes	no	yes	no	?
yes	no	yes	yes	?
yes	yes	no	no	yes
yes	yes	no	yes	yes
yes	yes	yes	no	?
yes	yes	yes	yes	?

<sup>a</sup> Rows with “?” in this column lack cases—the outcome cannot be determined.

with each combination. Table 31.2 illustrates a simple truth table with four dichotomous<sup>4</sup> conditions and sixteen combinations (called *configurations*).

In more complex truth tables the rows (combinations of conditions) may be quite numerous, for the number of configurations is a geometric function of the number of conditions (number of causal combinations =  $2^k$ , where  $k$  is the number of causal conditions). The use of truth tables to unravel causal complexity is described in detail elsewhere (e.g. Ragin 1987; Rihoux and Ragin 2008; Schneider and Wagemann forthcoming). The essential point is that the truth table elaborates and formalizes one of the two key analytic strategies of comparative research—examining cases sharing specific combinations of conditions to see if they share the same outcome. The goal of truth table analysis is thus to identify specific connections between combinations of conditions and outcomes.

### 3 THE STATE OF THE ART<sup>5</sup>

At present, several hundreds of QCA applications have been referenced worldwide, in several languages.<sup>6</sup> In terms of disciplinary orientation, more than two-thirds of the applications are found in political science (especially comparative politics and policy analysis; see Rihoux and Grimm 2006) and sociology. There is also a growing number of applications in other disciplines such as political economy, management studies, and criminology. Finally, some applications can be found in history, geography, psychology, education studies, and various other disciplines.

Although QCA is mainly designed for small- and intermediate- $N$  research, there is substantial variation across studies in the number of cases. Quite a few applications have a very small  $N$ , between three and six/seven cases. In the intermediate- $N$  range, most applications are to be found in the broad range from ten to fifty cases. However, several applications address much more cases, up to large- $N$  designs (e.g. Ishida, Yonetani, and Kosaka 2006).

The nature of the cases studied is also diverse. In most applications, cases (and outcomes) are macro- or meso-level phenomena, such as policy fields, collective actors, country or regional characteristics, and so on. There is also substantial variation as to the number of conditions included in the analysis, though of course there is (or at least there should be) some connection between the number of cases and the number of conditions (Rihoux and Ragin 2008). The vast majority of applications consider between three and six conditions—thus models elaborated for QCA analysis tend to be parsimonious.

<sup>4</sup> The procedures described here are *not* dependent on the use of dichotomies. Truth tables can be built from dichotomies (csQCA), multichotomies (mvQCA), and also from fuzzy sets (with set memberships in the interval from 0 to 1).

<sup>5</sup> For more details, see Rihoux (2006); Yamasaki and Rihoux (2008).

<sup>6</sup> A comprehensive list of applications is available through the COMPASS international bibliographical database: <<http://www.compass.org>>.

Last but not least, one particularly interesting development in QCA applications is the explicit combination with other types of methods, both qualitative and quantitative. Most often, there is already a lot of upstream qualitative work involved in the process of achieving an in-depth understanding of cases (Rihoux and Lobe 2008). With regards to the combination of QCA with other formal—mainly quantitative—methods, several fruitful attempts have been made to confront QCA with some more or less mainstream quantitative techniques such as discriminant analysis, factor analysis, various types of multiple regression, logistic regression, and logit regression.<sup>7</sup>

In terms of software, two major programs have been developed: FSQCA (for csQCA, fsQCA, and fuzzy sets) and TOSMANA (for csQCA and mvQCA). They offer complementary tools, in a quite user-friendly environment. In recent years, some additional tools have been developed. FSQCA now includes routines for truth table analysis of fuzzy-set data (see Ragin 2008, and below), calculations of consistency and coverage measures for both crisp and fuzzy-set analyses (Ragin 2006), and the possibility to derive three different solutions for each analysis: the more complex one, the most parsimonious one, and the intermediate one. TOSMANA now includes graphical aids such as the “visualizer” which produces Venn diagrams, and the “thresholdsetter” which gives a visual grip on the dichotomization or trichotomization thresholds.

#### 4 USES AND BEST PRACTICES

The QCA techniques can be used for at least five different purposes (De Meur and Rihoux 2002; Rihoux and Ragin 2008). (a) They may be used in a straightforward manner simply to summarize data in the form of a truth table, using it as a tool for data exploration. (b) Researchers may take advantage of QCA to check the coherence of their data, mainly through the detection of logical contradictions (i.e. similar combinations of conditions which, however, produce a different outcome in different cases). (c) They can be used to test hypotheses or existing theories. (d) Another use, quite close to the former, is the quick test of any assumption formulated by the researcher—that is, without testing a preexisting theory or model as a whole. This is another way of using QCA techniques for data exploration. (e) Finally, they may be used in the process of developing new theoretical assumptions in the form of hypotheses, following a more inductive approach. Researchers should determine which uses of QCA best suit their research goals. Some of these five uses of QCA are still very much underexploited.

In the last few years, some “best practices” have been more systematically laid out. Here we only discuss some main general guidelines (for more detailed technical

<sup>7</sup> For a list, see Rihoux and Ragin (2008). Noteworthy examples include Cronqvist and Berg-Schlosser (2006), Amenta and Poulsen (1996), and Dumont and Bäck (2006).

guidelines see Rihoux and Ragin 2008; Wagemann and Schneider 2007). First, it is advisable to draw on the different functions of the software. Many of these functions are still underused, such as the “hypothesis testing” function (e.g. Peillon 1996), which can be exploited in different ways. Second, technical and reference concepts should be used with care, to avoid confused and imprecise communication. Several misunderstandings—and misplaced critiques of QCA—stem from the misuse of technical terms. One of the most frequent examples is the reference to “independent variables” instead of “conditions” (see footnote 2).

Third, one should never forget the fundamentally configurational logic of QCA (Nomiya 2004). Hence one should never consider the influence of a condition in an isolated manner, especially in the interpretation of the solution of a truth table. Fourth, QCA should never be used in a mechanical manner, but instead as a tool that requires iterative steps (Rihoux 2003). With QCA, there are frequent moves back and forth between the QCA analysis proper (i.e. use of the software) and the cases, viewed in the light of theory. Bottom line: The use of QCA should be both case informed (relying on case-based knowledge: Ragin and Becker 1992) and theory informed. When researchers encounter difficulties, they should explain, as transparently as possible, how they have been resolved. This often implies being transparent about trade-offs, pragmatic choices which may at times seem somewhat arbitrary in real-life research. But at least the reader is informed about the choices that have been made and their rationale. Once again: The QCA programs should never be used in a push-button logic, but rather in a careful, self-conscious way. Needless to say, the same should go with any formal tool—e.g. statistical tools as well—in social science research.

Fifth, one should be careful in the interpretation of the solution of a truth table (called a minimal formula). In particular, it is advisable to be cautious before interpreting a minimal formula in terms of causality. Technically speaking, the formula expresses, more modestly, co-occurrences reflecting *potential* causal connections. It is then up to researchers to decide (relying on their substantive and theoretical knowledge) how far they can go in the interpretation in terms of “causality.” Finally, in the research process, it is almost always fruitful to use different methods. At different stages of empirical research, it is often the case that different methods suit different needs. Thus it is advisable to use QCA in some stages, while exploiting other methods (qualitative or quantitative) at other stages. This is not to say that QCA should necessarily be used in a modest way, as it can be used as the main data-analytic tool.

#### 5 BEYOND csQCA: mvQCA, FUZZY SETS, FSQCA, AND OTHER INNOVATIONS

Beyond the initial, dichotomous technique (csQCA), which remains the most used so far, two strands of techniques have been developed, allowing us to process more

fine-grained-data: mvQCA (multi-value QCA), on the one hand, and fuzzy sets on the other hand, with two variants: fuzzy sets proper, and fsQCA (fuzzy-set QCA), a procedure which links fuzzy sets to truth table analysis.

MvQCA has been developed along with the TOSMANA software (Cronqvist 2005; Cronqvist and Berg-Schlösser 2008). One problem in applying csQCA is the compulsory use of dichotomous conditions, which bears the risk of information loss and may create a large number of contradictory configurations (i.e. configurations with the same condition values, but yet leading to different outcome values). It can also lead to instances where two cases with somewhat different raw values are assigned the same Boolean value and/or two cases with quite similar values are assigned different Boolean values. As the name suggests, mvQCA is an extension of csQCA. It retains the main idea of csQCA, namely to perform a synthesis of a data-set, with the result that cases with the same outcome value are expressed by a parsimonious minimal formula. As in csQCA, the minimal formula contains one or several prime implicants, each of which covers a number of cases with this outcome, while no cases with a different outcome are explained by any of the prime implicants of that minimal formula. The key difference is that mvQCA also enables the inclusion of multi-value conditions. In fact, mvQCA is a generalization of csQCA, because indeed a dichotomous variable is a specific subtype of multi-value variables—it is simply a multi-value variable with only two possible values. In practical terms, the researcher can choose to use more than two values (typically trichotomies, but one can also use four categories or more) for some conditions, hence adding more diversity. MvQCA has already been applied successfully on different sorts of data-sets (e.g. Cronqvist and Berg-Schlösser 2006).

One apparent limitation of the truth table approach, which underlies both csQCA and mvQCA, is that it is designed for conditions that are simple presence/absence dichotomies (the Boolean, csQCA variant) or multichotomies (mvQCA). However, many of the conditions that interest social scientists vary by level or degree. For example, while it is clear that some countries are democracies and some are not, there is a broad range of in-between cases. These countries are not fully in the set of democracies, nor are they fully excluded from this set. Fortunately, there is a well-developed mathematical system for addressing partial membership in sets, fuzzy-set theory, which has been further elaborated for social scientific work by Ragin (2000). Fuzzy sets are especially powerful because they allow researchers to calibrate partial membership in sets using values in the interval between 0 (nonmembership) and 1 (full membership) without abandoning core set-theoretic principles such as, for example, the subset relation. As Ragin (2000) demonstrates, the subset relation is central to the analysis of causal complexity. In many respects fuzzy sets are simultaneously qualitative and quantitative, for they incorporate both kinds of distinctions in the calibration of degree of set membership. Thus, fuzzy sets have many of the virtues of conventional interval-scale variables, but at the same time they permit set-theoretic operations which are outside the scope of conventional variable-oriented analysis.

Fuzzy sets extend crisp sets by permitting membership scores in the interval between 0 and 1. For example, a country (e.g. the United States) might receive a

membership score of 1 in the set of rich countries but a score of only 0.9 in the set of democratic countries. The basic idea behind fuzzy sets is to permit the scaling of membership scores and thus allow partial or fuzzy membership. Thus fuzzy membership scores address the varying degree to which different cases belong to a set (including two qualitatively defined states: full membership and full nonmembership), as follows: A fuzzy membership score of 1 indicates full membership in a set; scores close to 1 (e.g. 0.8 or 0.9) indicate strong but not quite full membership in a set; scores less than 0.5 but greater than 0 (e.g. 0.2 and 0.3) indicate that objects are more “out” than “in” a set, but still weak members of the set; and finally a score of 0 indicates full nonmembership in the set.

Thus, fuzzy sets combine qualitative and quantitative assessment: 1 and 0 are qualitative assignments (“fully in” and “fully out,” respectively); values between 0 and 1 indicate partial membership. The 0.5 score is also qualitatively anchored, for it indicates the point of maximum ambiguity (fuzziness) in the assessment of whether a case is more “in” or “out” of a set. Note that fuzzy-set membership scores do not simply rank cases relative to each other. Rather, fuzzy sets pinpoint qualitative states while at the same time assessing varying degrees of membership between full inclusion and full exclusion. In this sense, a fuzzy set can be seen as a continuous variable that has been purposefully calibrated to indicate degree of membership in a well-defined set. Such calibration is possible only through the use of theoretical and substantive knowledge, which is essential to the specification of the three qualitative breakpoints: full membership (1), full nonmembership (0), and the crossover point, where there is maximum ambiguity regarding whether a case is more “in” or more “out” of a set (0.5) (Ragin 2008; this volume; Schneider and Wagemann forthcoming). Such calibration should not be mechanical—when specifying the qualitative anchors, the investigator should present a rationale for each breakpoint. In fact the qualitative anchors make it possible to distinguish between relevant and irrelevant variation, especially when one uses quantitative, interval-level data. For instance, variation in GNP per capita among the unambiguously rich countries is *not* relevant to membership in the set of rich countries, at least from the perspective of fuzzy sets. If a country is unambiguously rich, then it is accorded full membership, a score of 1. Similarly, variation in GNP per capita among the unambiguously not-rich countries is also irrelevant to degree of membership in the set of rich countries because these countries are uniformly and completely out of the set of rich countries. Thus, in research using fuzzy sets it is not enough simply to develop scales that show the relative positions of cases on distributions (e.g. a conventional index of wealth such as GNP per capita). Note, finally, that in a fuzzy-set analysis both the outcome and the conditions are represented using fuzzy sets.

However, a limitation of fuzzy sets is that they are not well suited for conventional truth table analysis. With fuzzy sets, there is no simple way to sort cases according to the combinations of conditions they display because each case’s array of membership scores may be unique. Ragin (2000) circumvents this limitation by developing an algorithm for analyzing configurations of fuzzy-set memberships that bypasses truth table analysis altogether. While this algorithm remains true to fuzzy-set theory



through its use of the containment (or inclusion) rule, it forfeits many of the analytic strengths that follow from analyzing evidence in terms of truth tables. For example, truth tables are very useful for investigating “limited diversity” and the consequences of different “simplifying assumptions” that follow from using different subsets of “logical remainders” to reduce complexity (see Ragin 1987; Ragin and Sonnett 2004; Rihoux and De Meur 2008). Analyses of this type are difficult when not using truth tables as the starting point.

Therefore, Ragin (2008) has built a bridge between fuzzy sets and truth tables, allowing one to construct a conventional Boolean truth table, starting from fuzzy-set data—this procedure is referred to as “fsQCA”. It is important to point out that this new technique takes full advantage of the gradations in set membership central to the constitution of fuzzy sets and is not predicated upon a dichotomization of fuzzy membership scores. Rather, the original interval-scale data is converted into fuzzy membership scores (which range from 0 to 1), thereby avoiding dichotomizing or trichotomizing the data (i.e. sorting the cases into crude categories). Actually, fsQCA offers a new way to conduct fuzzy-set analysis. This new analytic strategy is superior in several respects to the initial fuzzy-set strategy developed by Ragin (2000). While both approaches have strengths and weaknesses, the specificity of fsQCA is that it uses the truth table as the key analytic device. A further advantage of this procedure is that it is more transparent. Thus, the researcher has more direct control over the process of data analysis. This type of control is central to the practice of case-oriented research. Technically speaking, once a crisp truth table has been produced (by summarizing the results of multiple fuzzy set assessments), it is then analyzed using Boolean algebra.

Beyond the development of mvQCA, fuzzy sets, and fsQCA, quite a few specific innovations need to be mentioned. A first set of innovations deals with the issue of temporality, which indeed is not “built in” the QCA procedures—actually, from a case-oriented perspective, this is probably the main limitation of QCA so far. Some scholars have attempted to include the time dimension in applying the overall QCA technique, i.e. by specifying sequence considerations in the interpretation of the minimal formulae, by including dynamic considerations in the operationalization of conditions, or by segmenting cases following a chronological logic (e.g. Rihoux 2001; Clément 2005). Some first attempts have also been made to insert temporality into the analytic (software) procedure itself. Duckles, Hager, and Galaskiewicz (2005), using Event Structure Analysis (ESA), first construct some event structures, some of which are operationalized in sequential sub-models for successive QCA minimization procedures. Eventually they elaborate a complete model which enables them to identify some key precipitating factors in the chain of events, at least for some clusters of cases. Another attempt, by Caren and Panofsky (2005), consists in integrating temporality directly into QCA. Using an hypothetically constructed example, they argue that it is possible to develop an extension of QCA (TQCA—temporal QCA) to capture causal sequences. First, they include sequence considerations as a specific case attribute, hence increasing dramatically the number of possible configurations. Second, they place theoretical restrictions to limit the number of possible configurations. Third, they perform a specific form of Boolean minimization, which yields richer minimal

formulae which also include sequences and trajectories. In a somewhat other direction, Krook (2006) has made a first attempt to link QCA procedures with an “optimal matching” procedure.

One should also mention the “two-step” procedure initiated by Schneider and Wagemann (2006; forthcoming), which draws a distinction between remote and proximate conditions. Along the same line of enriching and refining the basic QCA procedures, a series of complementary tests have been developed—specifically: testing for the presence of necessary or sufficient conditions, and also assessing the “consistency” and “coverage” of the different elements of the minimal formula (Ragin 2006; Schneider and Wagemann forthcoming). Another adjacent development is the systematization of a procedure (called MSDO/MDSO) which can be used to boil down the number of potential conditions when they are too numerous, while also shedding some light on how cases cluster, before engaging in QCA proper (Berg-Schlosser and De Meur 1997; 2008). Within the QCA procedures themselves, several refinements have also been successfully applied, such as a more informed use of “logical remainders” (nonobserved cases), a specific treatment to solve the so-called “contradictory simplifying assumptions” (“hidden” contradictions which might occur when one uses logical remainders), and so on. The good point here is that these refinements all go in the direction of further enhancing the transparency and solidity of the analysis, and that they have been translated into good practices.

## 6 CONCLUSION: PROMISES AND OPENINGS

The audience for QCA and its techniques is growing; as are debates about its strengths and limitations. A detailed discussion thereof lies beyond the scope of this chapter (for a full view, see De Meur, Rihoux, and Yamasaki 2008; Ragin and Rihoux 2004<sup>8</sup>). Some key topics of discussion include dichotomization, the use of nonobserved cases (the “logical remainders”), case sensitivity, model specification, causality, and temporality. The bottom line is that many of the critiques are misplaced or overstated, especially those which are formulated from a mainstream quantitative (i.e. statistical) perspective, because they fail to grasp the specificity of the configurational, case-oriented foundation of QCA and its techniques. Of course, there are limitations, as with any technique. It is precisely some of these limitations which foster the development of new tools and ways to enrich QCA.

Quite a few innovations are expected, in the short to medium term, or are already being initiated at present. Some further software innovations will surely be developed within the next few years. Some other efforts are also being undertaken on other platforms, such as the R software (Dusa 2007). Here are some issues on the agenda, which

<sup>8</sup> See also issue 2.2. (2004) of the *Qualitative Methods Newsletter* (APSA), as well as issue 40.1. (2005) of *Studies in Comparative International Development*.

will hopefully materialize at some stage in the software developments, through FSQCA, TOSMANA, or other platforms: a more explicit inclusion of the time dimension in the computing procedures; some further improvements in the user-friendliness of the platforms; some interconnections with other software (e.g. importing/exporting data); new ways to visualize the configurations as well as the minimal formulae, etc.

Another particularly promising path consists in small- or intermediate-*N* research designs in which cases are individuals (micro-level cases). Especially in more participatory research designs, i.e. when researchers are able to engage in regular interaction with the individuals (the “cases”) being the object of the study, one may argue that they attain an even better understanding of each individual case than would be the case for meso- or macro-level phenomena. Indeed, they are literally able to interact directly with each and every one of the cases, which would prove much more difficult when cases are meso- or macro-level phenomena (Lobe and Rihoux 2008).

QCA and linked techniques are still quite new tools. They display broad potential in terms of discipline, number of cases, research design, and types of uses. Rather than being a middle path between case-oriented and variable-oriented research, as Ragin (1987) initially argued, all things considered, QCA is more related to case-study methods—especially the crisp, dichotomous version (Rihoux and Lobe 2008). As with any case-oriented methods, a researcher using QCA meets a trade-off between the goals of reaching a certain level of theoretical parsimony, establishing explanatory richness, and keeping the number of cases at a manageable level (George and Bennett 2005). What is specific about QCA, as compared with “focused, structured comparison,” is that it enables one to consider a larger number of cases, provided one is willing to accept a certain level of simplification and synthesis necessitated for Boolean or set treatment. Still, in that process, one does not sacrifice explanatory richness, and the possible generalizations which will be produced will always be contingent, in the sense that they only apply to some specific types or clusters of well-delineated cases which operate in specific contexts (George and Bennett 2005).

Naturally, QCA and connected techniques should not be viewed in isolation. They are compatible with other approaches, especially comparative historical analysis (Mahoney and Rueschemeyer 2003; Mahoney and Terrie, this volume) and theory-led case-oriented research (George and Bennett 2005; Bennett, this volume; Gerring 2006; this volume). In addition, much progress can be expected within the next few years, when QCA and connected techniques will hopefully be combined/confronted more systematically with other techniques, be they more qualitative or more quantitative. At the same time, we can also expect some significant further development in terms of software (see above). In addition, dissemination of these techniques is now being extended through some training programs and specialized courses in various institutions. Among other dissemination efforts, a first overarching English-language textbook (Rihoux and Ragin 2008) will make these techniques more accessible to students. Of course neither csQCA, nor mvQCA, fuzzy sets, or fsQCA, solve all problems faced by empirical researchers—no technique should be expected to accomplish this. Yet the application of these techniques is of great analytic value and has potential

in many fields across political science and social sciences broadly defined, and in different research designs.

## REFERENCES

- AMENTA, E., and POULSEN, J. D. 1996. Social politics in context: the institutional politics theory and social spending at the end of the New Deal. *Social Forces*, 75: 33–60.
- BERG-SCHLOSSER, D., and DE MEUR, G. 1997. Reduction of complexity for a small-n analysis: a stepwise multi-methodological approach. *Comparative Social Research*, 16: 133–62.
- . 2008. Comparative research design: case and variable selection. In *Configurational Comparative Methods*, ed. B. Rihoux and C. Ragin. Thousand Oaks, Calif.: Sage.
- CAREN, N., and PANOFSKY, A. 2005. TQCA: a technique for adding temporality to qualitative comparative analysis. *Sociological Methods and Research*, 34: 147–72.
- CLÉMENT, C. 2005. The nuts and bolts of state collapse: common causes and different patterns? COMPASS Working Paper 2005–32.
- CRONQVIST, L. 2005. Introduction to multi-value qualitative comparative analysis (MVQCA). COMPASS Didactics Paper No. 4.
- and BERG-SCHLOSSER, D. 2006. Determining the conditions of HIV/AIDS prevalence in sub-Saharan Africa: employing new tools of macro-qualitative analysis. Pp. 145–66 in *Innovative Comparative Methods for Policy Analysis*, ed. B. Rihoux and H. Grimm. New York: Springer.
- . 2008. Multi-value QCA (mvQCA). In *Configurational Comparative Methods*, ed. B. Rihoux and C. Ragin. Thousand Oaks, Calif.: Sage.
- DE MEUR, G., and RIHOUX, B. 2002. *L'Analyse quali-quantitative comparée (AQQC-QCA): approche, techniques et applications en sciences humaines*. Louvain-la-Neuve: Academia-Bruylant.
- RIHOUX, B., and YAMASAKI, S. 2008. Addressing the critiques of QCA. In *Configurational Comparative Methods*, ed. B. Rihoux and C. Ragin. Thousand Oaks, Calif.: Sage.
- DUCKLES, B. M., HAGER, M. A., and GALASKIEWICZ, J. 2005. How nonprofits close: using narratives to study organizational processes. Pp. 169–203 in *Advances in Qualitative Organizational Research*, ed. K. D. Elsbach. Greenwich, Conn.: Information Age.
- DUMONT, P., and BÄCK, H. 2006. Why so few and why so late? Green parties and the question of governmental participation. *European Journal of Political Research*, 45: s35–68.
- DUSA, A. 2007. User manual for the QCA (GUI) package in R. *Journal of Business Research*, 60: 576–86.
- GEORGE, A. L., and BENNETT, A. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass.: MIT Press.
- GERRING, J. 2006. *Case Study Research: Principles and Practice*. Cambridge: Cambridge University Press.
- GOERTZ, G., and STARR, H. 2002. *Necessary Conditions: Theory, Methodology, and Applications*. New York: Rowman and Littlefield.
- ISHIDA, A., YONETANI, M., and KOSAKA, K. 2006. Determinants of linguistic human rights movements: an analysis of multiple causation of LHRs movements using a Boolean approach. *Social Forces*, 84: 1937–55.
- KROOK, M. L. 2006. Temporality and causal configurations: combining sequence analysis and fuzzy set/qualitative comparative analysis. Presented at the annual APSA meeting, Philadelphia.

- LOBE, B., and RIHOUX, B. 2008. The added value of micro-level QCA: getting more out of rich case knowledge. Unpublished manuscript.
- MAHONEY, J., and RUESCHEMEYER, D. 2003. *Comparative Historical Research*. Cambridge: Cambridge University Press.
- NOMIYA, D. 2004. Atteindre la connaissance configurationnelle: remarques sur l'utilisation précautionneuse de l'AQCC. *Revue Internationale de Politique Comparée*, 11: 131–3.
- PEILLON, M. 1996. A qualitative comparative analysis of welfare legitimacy. *Journal of European Social Policy*, 6: 175–90.
- RAGIN, C. C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- 2006. Set relations in social research: evaluating their consistency and coverage. *Political Analysis*, 14: 291–310.
- 2008. Qualitative Comparative Analysis using fuzzy sets (fsQCA). In *Configurational Comparative Methods*, ed. B. Rihoux and C. Ragin. Thousand Oaks, Calif.: Sage.
- and BECKER, H. 1992. *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press.
- and RIHOUX, B. 2004. Replies to commentators: reassurances and rebuttals. *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, 2: 21–4.
- and SONNETT, J. 2004. Between complexity and parsimony: limited diversity, counterfactual cases, and comparative analysis. COMPASSS Working Paper 2004–23.
- RIHOUX, B. 2001. *Les Partis politiques: organisations en changement. Le test des écologistes*. Paris: L'Harmattan.
- 2003. Bridging the gap between the qualitative and quantitative worlds? A retrospective and prospective view on qualitative comparative analysis. *Field Methods*, 15: 351–65.
- 2006. Qualitative comparative analysis (QCA) and related systematic comparative methods: recent advances and remaining challenges for social science research. *International Sociology*, 21: 679–706.
- and GRIMM, H. (eds.) 2006. *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative–Qualitative Divide*. New York: Springer/Kluwer.
- and LOBE, B. 2008. The case for QCA: adding leverage for thick cross-case comparison. In *Handbook of Case Study Methods*, ed. C. C. Ragin and D. Byrne. Thousand Oaks, Calif.: Sage.
- and RAGIN, C. (eds.) 2008. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Applied Social Research Methods. Thousand Oaks, Calif.: Sage.
- SCHNEIDER, C. Q., and WAGEMANN, C. 2006. Reducing complexity in qualitative comparative analysis (QCA): remote and proximate factors and the consolidation of democracy. *European Journal of Political Research*, 45: 751–86.
- forthcoming. *Qualitative Comparative Analysis (QCA) and Fuzzy Sets: A User's Guide*. German-language version (2007), *Qualitative Comparative Analysis (QCA) und Fuzzy Sets*. Opladen: Verlag Barbara Budrich.
- WAGEMANN, C., and SCHNEIDER, C. Q. 2007. Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. COMPASSS Working Paper 2007–51.
- YAMASAKI, S., and RIHOUX, B. 2008. A commented review of applications. In *Configurational Comparative Methods*, ed. B. Rihoux and C. Ragin. Thousand Oaks, Calif.: Sage.

## CHAPTER 32

---

COMPARATIVE-  
HISTORICAL  
ANALYSIS IN  
CONTEMPORARY  
POLITICAL  
SCIENCE

---

JAMES MAHONEY  
P. LARKIN TERRIE

ALTHOUGH comparative-historical analysis has roots as far back as the founders of modern social science, its place in contemporary political science can be traced to a series of successful books published in the 1960s and 1970s, such as Moore (1966), Bendix (1974), Lipset and Rokkan (1968), Tilly (1975), and Skocpol (1979).

Over the last twenty years, the tradition has sustained momentum in part through the publication of scores of major new books. This scholarship includes work across many of the key substantive areas of comparative politics: social provision and welfare state development (e.g. Esping-Anderson 1990; Hicks 1999; Huber and Stephens 2001; Pierson 1994; Skocpol 1992; Steinmo 1993); state formation and state restructuring (Bensel 1990; Ekiert 1996; Ertman 1997; Tilly 1990; Waldner 1999); economic development and market-oriented adjustment (Bunce 1999; Evans 1995; Haggard 1990; Karl 1997; Kohli 2004; Sikkink 1991); racial, ethnic, and national identities (Lustick

1993; Marx 1998; Yashar 2005); revolutionary change (e.g. Goldstone 1991; Goodwin 2001; Wickham-Crowley 1992); and democratic and authoritarian regimes (Collier 1999; Collier and Collier 1991; Downing 1992; Linz and Stepan 1996; Luebbert 1991; Mahoney 2001; Rueschemeyer, Stephens, and Stephens 1992).

The comparative-historical approach has been strongly identified with the publication of books. Yet beyond these books, key elements of this tradition are also found in a significant portion of published work in scholarly journals on comparative politics—a fact that we empirically demonstrate in this chapter.

While comparative-historical analysis “has claimed its proud place as one of the most fruitful research approaches in modern social science” (Skocpol 2003, 424), it is also true that methodological aspects of the approach are still received skeptically in some quarters. Perhaps most notably, scholars who pursue the statistical testing of hypotheses with large numbers of cases have raised concerns about this tradition. They have argued that, from the standpoint of statistical methodology, this line of enquiry violates well-known aspects of good research design and procedure (e.g. Coppedge 2008; King, Keohane, and Verba 1994; Geddes 1990; 2003; Goldthorpe 1997; Lieberman 1991; 1994; 1998). They use these criticisms as a basis for questioning whether the influential substantive findings produced in this field are, in fact, valid.

In this chapter, we suggest that existing concerns arise from a fundamental misunderstanding of the goals and methods of comparative-historical analysis. This misunderstanding, in turn, is linked to a failure to appreciate basic differences between comparative-historical and statistical analysis. We show that these two research traditions are best understood as adopting distinct research goals, using different methods to achieve these goals, and thus quite justifiably pursuing different kinds of overall research designs. Once basic differences in research orientations are recognized, it becomes clear that advice and criticisms derived solely from a statistical template are not appropriate (see also Brady and Collier 2004; Mahoney and Goertz 2006).

Clarifying the differences between comparative-historical and statistical analysis helps to promote a more fruitful dialogue among scholars. Despite their different research objectives and contrasting methodological tools, researchers representing these different traditions stand to benefit from better understanding one another's methods and research practices. There are at least two reasons why. First, insights from one tradition often can stimulate new and useful ideas for the other tradition. For example, insights about combinatorial causation and equifinality in comparative-historical methods have led to the creation of new statistical methods (Braumoeller 2003; 2004). The same is true of recent writings on necessary and sufficient causes (Clark, Gilligan, and Golder 2006). Likewise, comparing frequentist and Bayesian statistics has stimulated new insights about process tracing in comparative-historical research (Bennett 2006). And statistical techniques have been combined with formal qualitative comparative analysis in creative ways (Ragin 2000).

Second, the proliferation of multimethod research in contemporary political science makes knowledge of a wide range of methods increasingly important. Obviously, scholars who themselves pursue multimethod research should be well schooled in all of the relevant methodological traditions. At the same time, it seems increasingly

important that methodologists themselves be able to offer sound advice to scholars who seek to combine statistical and case-study methods, including comparative-historical methods. Obviously, no one methodologist can be expected to be an expert across the board; however, methodologists should know when the limits of their expertise are reached and thus when it is time to defer to specialists in other methodological orientations.

## 1 THE FIELD OF COMPARATIVE-HISTORICAL ANALYSIS

Many famous books are strongly associated with the comparative-historical tradition. But one might reasonably wonder about the overall commonality of the approach in contemporary political science, including in journals. Does a significant body of literature exist beyond the famous examples? How could we identify such a literature if it did? We address these questions by measuring several traits associated with this tradition and assessing empirically the extent to which these traits are found together in published studies on comparative politics.

As with any research orientation, there are different ways of defining comparative-historical analysis. According to Mahoney and Rueschemeyer (2003), this approach investigates “big questions”—substantively important and large-scale outcomes—that take the form of puzzles about specific cases. In addressing these puzzles, researchers are centrally concerned with causal analysis, the examination of processes over time, and use of systematic and contextualized comparison. This understanding of the field is similar to that adopted by Collier (1998) and Skocpol (1979, 36–7; 1984, 1).

In this chapter, we are especially interested in defining the field in terms of characteristics that can be empirically measured. With this in mind, we emphasize here three core traits and two secondary traits as features that are important to most work in the field. The three core traits concern explanatory goal, conception of causation, and method of theory testing. On each of these three dimensions, comparative-historical analysis directly contrasts with statistical analysis (see Tables 32.1 and 32.2). Comparative-historical work adopts a causes-of-effects approach to explanation, a necessary and/or sufficient conception of causation, and process tracing to test theories. By contrast, statistical analysis uses an effects-of-causes approach to explanation, an average effects conception of causation, and regression techniques to test theories. In addition to these three main dimensions, two other attributes are often associated with comparative-historical work: the use of a comparative set-theoretic logic and the analysis of temporal sequencing and/or path dependence.<sup>1</sup> In Appendix A, we discuss the definition and measurement of each of these traits.

<sup>1</sup> We also coded studies according to whether they adopt a rational choice framework.

**Table 32.1.** Attributes associated with comparative-historical analysis

Attribute	Definition
Causes-of-effects approach <sup>a</sup>	Research goal is to provide complete explanations of specific outcomes in particular cases.
Necessary/sufficient conception of causation <sup>a</sup>	Study treats individual causal factors or sets of multiple causal factors as necessary/sufficient for the outcome of interest.
Process-tracing method <sup>a</sup>	Study explores the mechanisms, within particular cases, through which potential causal factors are hypothesized to affect an outcome.
Comparative set-theoretic methods	Study tests theory using set-theoretic methods (e.g. Mill's methods of agreement and difference, Boolean algebra).
Temporal processes modeled	Explanation emphasizes the sequencing of independent variables and/or path-dependent processes of change.

<sup>a</sup> Core attribute of comparative-historical analysis.

**Table 32.2.** Attributes associated with statistical analysis

Attribute	Definition
Effects-of-causes approach	Research goal is to estimate the effects of one or more independent variables on a dependent variable/s across a large number of cases.
Average effects conception of causation	Study treats independent variables as parameters whose average effects can be estimated across the full population of cases.
Regression methods	Regressions are used for theory testing.

With these defining traits at hand, we explored empirically whether a tradition of comparative-historical analysis could be found within the subfield of comparative politics. We gathered data from articles that recently appeared in the major comparative politics journals—*Comparative Political Studies*, *Comparative Politics*, and *World Politics*. We set out to sample approximately 100 articles, evenly distributed, from these journals.<sup>2</sup> To make the sample representative of recent work in comparative politics, we first coded articles from 2005 for each journal. Articles from earlier years were coded if doing so was necessary to obtain a sufficiently large sample for the journal. In order to check the robustness of our results, we also analyzed approximately forty articles on comparative politics from the main discipline-wide journals—the *American Journal of Political Science*, the *American Political Science Review*, and the *Journal of Politics*.<sup>3</sup>

<sup>2</sup> The final sample consisted of 107 articles: 30 from *CP*, 38 from *CPS*, and 39 from *WP*. Note that descriptive, theoretical, and methodological articles were excluded from the sample.

<sup>3</sup> This sample consisted of 42 articles: 13 from *AJPS*, 15 from *APSR*, and 14 from *JOP*. These journals were not included in the original sample because the empirical studies they publish are almost exclusively statistical, and as such, they are less representative than the subfield journals of the

**Table 32.3.** Frequency of attributes associated with comparative-historical analysis

CHA attribute	Percentage of articles
Causes-of-effects approach	55.1
Necessary/sufficient conception of causation	57.9
Process-tracing method	58.9
Comparative set-theoretic methods	43.9
Temporal processes modeled	21.5

Note: Percentage of studies from sample of comparative politics journals ( $N = 107$ ).

The data reveal that all five of the attributes associated with the comparative-historical tradition commonly appear in the comparative politics literature, particularly in the articles published in the subfield journals. Table 32.3 reports the frequency with which each of these attributes appeared in the subfield journals. The three core traits of the field—a causes-of-effects approach to explanation, a necessary/sufficient conception of causation, and a process-tracing methodology—each appear in over half of these articles. The two secondary traits—a comparative set-theoretic logic and a concern with temporal sequencing or path dependence—also appear in a significant proportion of articles. What is more, these attributes are not randomly distributed among the articles in the sample. They have a marked tendency to cluster together, as the factor analysis results in Table 32.4 show. The three core attributes of the comparative-historical field are especially likely to hang together (as are the core attributes of statistical work). The temporal process/path dependence and comparative set-theoretic variables also tend to cluster with these three core attributes, though they exhibit lower factor loadings.

Given the high frequency of each individual attribute and the strength of their tendency to cluster, it is clear that a body of work that can be called comparative-historical analysis is relatively common in the literature on comparative politics. Table 32.5 displays the number of articles from the subfield journal sample that can be classified as comparative-historical analysis according to four possible definitions (articles that have any of the three core attributes associated with statistical work are excluded). Just under half of the articles in this sample have all three of the core attributes. When more restrictive definitions are used, the proportion declines, but a nontrivial percentage of articles still qualify as comparative-historical analysis.

Comparative-historical analysis is, in short, a leading research tradition in the subfield of comparative politics based on prominence of current usage alone. Given this prominent place, it seems quite important that we assess soberly the validity of methodological concerns that have been raised about the tradition.

methodological diversity in comparative politics (Mahoney 2007). Nevertheless, the inclusion of these additional articles did not significantly alter the initial findings. Hence, the factor analysis results presented in Table 32.4 are for the combined sample of articles from both the subfield and discipline-wide journals.

Table 32.4. Factor analysis of methodological attributes

	Factor 1
Causes-of-effects approach	0.9473
Necessary/sufficient conception of causation	0.9428
Process-tracing method	0.8854
Comparative set-theoretic methods	0.7169
Temporal processes modeled	0.4644
Effects-of-causes approach	-0.8654
Average effects conception of causation	-0.9321
Regression methods	-0.9149
Rational choice theory	-0.4490

Note: Results are from principal factors analysis performed on the entire sample of articles ( $N = 149$ ) from comparative politics and discipline journals. Reported is the first factor extracted, which explains 97.48 percent of the total variance in these nine variables (eigenvalue = 5.96).

## 2 CONCERNS ABOUT METHODOLOGICAL PRACTICES

All observational studies in the social sciences confront important obstacles and potentially are subject to error. However, some analysts have argued that comparative-historical research faces especially grave problems that can be avoided in statistical research (for a recent statement, see Coppedge 2008). The implicit or explicit implication is often that social scientists should pursue statistical research when possible (see also Lijphart 1971). In this section, by contrast, we argue that comparative-historical

Table 32.5. Frequency of comparative-historical articles

Definition of CHA	Percentage of articles
Causes-of-effects * necessary/sufficient * process tracing	49.5
Causes-of-effects * necessary/sufficient * process tracing * comparative set-theoretic methods	36.4
Causes-of-effects * necessary/sufficient * process tracing * temporal processes	18.7
Causes-of-effects * necessary/sufficient * process tracing * comparative set-theoretic methods * temporal processes	14.0

Note: Percentage of studies in sample from comparative politics journals ( $N = 107$ ). The "\*" symbol denotes the logical AND.

analysis and statistical analysis pursue different research goals, and that while they both face methodological challenges, they both play an essential role in generating knowledge in political science.

### 2.1 Selection Bias

Several methodologists have sounded alarm bells about the tendency of qualitative researchers to select cases based on their value on the dependent variable (Achen and Snidal 1989; Geddes 1991; 2003; King, Keohane, and Verba 1994). These stern warnings about deliberately selecting cases because they exhibit certain outcomes are especially applicable to comparative-historical studies, which quite explicitly engage in the practice. On the one hand, of course, selection on the dependent variable in this field is hardly surprising, given that the research goal is precisely the explanation of particular outcomes. If one wishes to explain certain outcomes, it seems natural to choose cases that exhibit those outcomes. On the other hand, however, selecting cases based on their value on the dependent variable will bias findings in statistical research. From this standpoint, the practice seems to violate a basic norm of good research.

To evaluate this concern, when applied to comparative-historical analysis, we need to recognize that the statistical literature on bias deriving from selection on the dependent variable assumes that one wishes to generalize about average causal effects from a sample to a well-defined larger population. In comparative-historical research, however, one seeks to identify realized causal effects in particular cases; generalizing about averages from a sample to a larger population is at most a secondary goal. Insofar as comparative-historical researchers select what can be considered the entire universe of cases, therefore, standard issues of selection bias do not arise, regardless of whether the cases were chosen for their values on the dependent variable (for more extensive discussions, see Collier and Mahoney 1996; Collier, Mahoney, and Seawright 2004).<sup>4</sup>

### 2.2 Scope and Generalization

These observations raise questions about generalization in comparative-historical analysis. Obviously, comparative-historical researchers cannot simply make up whatever definition of the universe of cases they so choose; the decision to limit a theory's applicability to a particular set of cases should not be arbitrary. One needs to ask, therefore, about the methodological basis for adopting a restrictive understanding of scope. This question is critical because the practice of restricting generalizations to a limited set of cases—not issues of selection bias, as conventionally understood—is

<sup>4</sup> Problems of selection bias, as conventionally understood in the statistical literature, arise in comparative-historical studies primarily when analysts seek to generalize their theories beyond the initial cases investigated.

often the real source of concern held by statistical methodologists about case selection in comparative-historical analysis. In particular, given the expansive definition of scope often employed in statistical research, the findings of comparative-historical researchers appear to be derived from a potentially unrepresentative sample of cases that is arbitrarily treated as the full population. Let us then explore these issues of scope and generalization.

Social scientists commonly impose scope restrictions on their findings to avoid problems associated with causal heterogeneity, which generates instability in estimates of causal effects. Indeed, primarily because of causal heterogeneity, social scientists of all traditions rarely develop theories that are intended to apply to all places and times. In addition to issues of causal heterogeneity, the need for stable concepts and measurement leads to the use of scope restrictions in which the analyst excludes cases where conceptual and measurement validity cannot be maintained.

In comparative-historical research, analysts adopt a narrow scope because they believe that causal and conceptual heterogeneity are the norm for their theories when assessed across large populations (Mahoney and Rueschemeyer 2003). But is this belief justified? Here we need to recognize that causal heterogeneity is not an ontological property inherent in a population of cases, but rather a feature of the *relationship* between a specific theory and a population of cases (Seawright and Collier 2004, 276; Goertz and Mahoney 2007). A given population of cases may be heterogeneous vis-à-vis one theory but not another. The same is true of conceptual heterogeneity: Cases may be heterogeneous vis-à-vis some concepts but not others. One key implication is that some types of theories (or concepts) may be more likely than others to produce heterogeneity as the size of the population of cases increases.

There are very good reasons for believing that the type of theories evaluated in comparative-historical analysis is especially likely to generate causal heterogeneity in response to even modest increases in population size. To understand why, we need to compare the problem of missing variables in comparative-historical analysis and statistical analysis. This discussion, in turn, will take us back to the contrasting research goals of the two traditions.

In this tradition, the exclusion of one or more important explanatory variables from a theory is appropriately regarded as a major problem (Ragin 2004, 135–8). This is true because the very goal of this kind of work is to explain particular outcomes in specific cases as completely and adequately as possible. All relevant evidence pertaining to the cases should be gathered and assessed. If theories are missing key variables, or have misspecified key relationships among the variables that are included, these facts count significantly against the arguments posited by the researcher. The failure of previous investigators to consider one or more critical variables in fact provides a common basis for comparative-historical analysts to criticize existing work and build new explanatory theories. Missing variables are thus a constant potential source of causal heterogeneity in this field. Similar arguments can be extended to measurement error, which needs to be addressed and eliminated completely for each specific case, if possible (Ragin 2004). Otherwise, the goal of adequately explaining an outcome in particular cases is compromised. In comparative-historical analysis,

indeed, theory falsification often occurs with the change in value of one or a small number of variables. Accordingly, in this mode of research, one needs to strive to avoid measurement error for the cases analyzed, otherwise conceptual heterogeneity problems will likely arise.

There are thus good methodological reasons related to the need to avoid causal and conceptual heterogeneity that explain why comparative-historical researchers restrict the scope of their analysis to a limited number of cases. Given the kind of explanatory theory that these analysts pursue, built around the idea of realized causal effects for particular outcomes, they must quite carefully and deliberately define their population of cases to try to avoid all heterogeneity problems. Once the population is defined, even a modest increase in the number of cases runs the risk of excluding key causal factors relevant to the new cases or introducing measurement problems for the variables that are already included in the theory. Because significant modifications to the theoretical model are often required as new cases are added, the best solution may be to impose restrictive scope conditions that limit generalization.

In statistical analysis, by contrast, the goal of research is typically to estimate the average effects of one or more independent variables. Given this goal, missing variables are not necessarily a problem as long as key assumptions, especially that of conditional independence, still hold. Independent variables that are important for only a small subset of cases may be appropriately considered “unsystematic” and relegated to the error term of a regression model. Indeed, even missing independent variables that are systematically related to the outcome of interest will not necessarily bias estimates of the average effects, as long as conditional independence still applies. Likewise, measurement error in statistical analysis does not raise the kinds of problem that it does in comparative-historical analysis. With large numbers of cases, measurement error is always present and cannot be completely eliminated. However, if one seeks to identify average effects, as statistical analysts do, measurement error is not a devastating problem as long as it is nonsystematic or at least can be adequately modeled in the event that it is systematic. Unbiased estimates of average effects are possible in the presence of measurement error.

The fact that statistical analysis can maintain causal homogeneity even in the presence of missing variables and measurement error allows this kind of research to embrace a more expansive understanding of scope and generalization than the comparative-historical approach. For example, the inclusion of new cases with outcomes that were partially caused by idiosyncratic factors will not necessarily raise any special heterogeneity problems in statistical analysis. As long as assumptions such as conditional independence are valid and measurement error can be modeled, the extension of the scope to include new cases is usually not a problem in statistical research. Not surprisingly, therefore, statistical researchers worry less about issues of heterogeneity as they extend their arguments to new cases.

An important issue arises at this point: If comparative-historical explanations are fragile when new cases are introduced, but statistical explanations are less fragile, does it not also follow that statistical explanations are “superior?” There are two reasons why this conclusion is not correct. The first is that the ability of statistical

analysis to adopt a wide scope of generalization is dependent on the validity of key assumptions, especially conditional independence. In contemporary political science, many empirical researchers feel quite comfortable making this assumption with little elaboration. Yet methodologists and statisticians often suggest that the assumption is an unrealistic leap of faith in much of the observational research pursued in the social sciences (Liebersohn 1985; Freedman 1991). Insofar as the assumption of conditional independence cannot be sustained, statistical research suffers from unrecognized and unmodeled causal heterogeneity. In other words, it is possible that the expansive understanding of scope adopted in statistical analysis is often not appropriate.

Second, and more important for our purposes, it is essential to remember that comparative-historical researchers and statistical researchers have distinct research goals. If one wishes to explain particular outcomes in specific cases, as comparative-historical researchers do, then one must formulate theories in which it is not possible to easily extend the scope of generalization. The alternative is to reject the research goals of comparative-historical work; that is, to prohibit studies that seek to explain particular outcomes in specific cases and encourage scholars only to ask questions about average effects across large populations. For reasons that we discuss below, this kind of prohibition against asking comparative-historical questions would be extremely costly for social science knowledge. In short, if one is going to remain open to different forms of knowledge accumulation, and allow scholars to ask such questions, then one must be willing to live with the restricted scope that accompanies comparative-historical analysis.

### 2.3 Assessing Causation with a Small *N*

Even if the limited scope of comparative-historical enquiry makes good sense, some analysts are still concerned that the small number of cases that fall within this scope does not permit the scientific testing of hypotheses. From a statistical standpoint, a small population poses a degrees of freedom problem and insurmountable obstacles for hypothesis testing. How can researchers ever hope to adjudicate among rival explanations if they select so few cases?

The answer to this question again requires appreciating differences between statistical analysis and comparative-historical analysis. In statistical research, where the goal is to estimate average causal effects, one needs to have enough cases to control for relevant variables and still achieve specified confidence levels. However, with comparative-historical research, the goal is not to generalize about typical effects for a large population. Rather, the goal is to determine whether a given variable *did exert* a causal effect on an outcome in a particular set of cases. Given this goal, researchers need to embrace a distinct understanding of causation and indeed of explanation, which—as we shall now see—obviates the need for a large number of cases to achieve valid causal assessment.

Comparative-historical researchers ask the following question about any potential causal factor: Did it exert an effect (alone or in combination with other variables)

on the specific outcomes of interest in the particular set of cases that comprise the population? Sometimes, even a cursory examination will allow one to dispose of certain causal factors that might be generally relevant across a large population of cases. For example, when explaining the emergence of democracy in economically poor India or Costa Rica, the variable of development is clearly not useful (at least not in the usual way), even though it is positively related to democracy in a large sample of cases. In other instances, however, plausible causal factors cannot be so quickly dismissed. Many potential causal factors are “correlated” with the specific outcome of interest. How do researchers adjudicate among these rival explanations that are matched with the outcome of interest?

Researchers in the comparative-historical tradition use the method of process tracing—which involves marshalling “within-case” data—to pass judgement on the validity of rival explanations emphasizing factors that cannot be eliminated through comparative matching techniques. Although here is not the place to discuss at length the mechanics of process tracing (see George and Bennett 2005), a few words are in order. Most basically, process tracing helps one to assess whether a posited causal factor actually exerts a causal effect on a specific outcome. This is done by exploring the mechanisms through which the potential causal factor is hypothesized to contribute to the outcome. If intervening mechanisms cannot be located, then doubt is cast upon the causal efficacy of the factor in question. By contrast, if appropriate intervening mechanisms are found, then one has grounds for believing that the factor in question did exert the effect. Beyond this, process tracing allows one to evaluate hypotheses by considering “sub-hypotheses” that do not necessarily refer to intervening mechanisms but that should be true if the main hypothesis of interest is valid (Mahoney and Villegas 2007).

It bears emphasis that this mode of hypothesis assessment does not require a large number of cases. Rather, like a detective solving a crime, the comparative-historical researcher who uses process tracing draws on particularly important facts from individual cases (see Goldstone 1997; McKeown 1999). Not all pieces of evidence count equally. Some forms of evidence are “smoking guns” that strongly suggest a theory is correct; others are “air-tight alibis” that strongly suggest a theory is not correct (Collier, Brady, and Seawright 2004). For these researchers, a theory is often only one key observation away from being falsified. Yet they may have certain kinds of evidence that suggest that the likelihood of theory falsification ever occurring is small.

Another relevant consideration concerns the conception of causation that is used in comparative-historical explanation. The various small-*N* comparative methods adopted by these researchers—Mill’s methods of agreement and difference, explanatory typologies, and qualitative comparative methods—all assume understandings of causation built around necessary and/or sufficient causes (Ragin 1987; 2000; Mahoney 2000; Goertz and Starr 2003; Elman 2005; George and Bennett 2005).<sup>5</sup> By contrast,

<sup>5</sup> Quite often, researchers treat individual causes as parts of a larger combination of causes that are together *jointly sufficient* for the outcome of interest (Mackie 1980). In fact, in this field, distinct



mainstream statistical methods assume forms of symmetrical causation that are not consistent with necessary and/or sufficient causation.

To assess hypotheses about necessary and sufficient causes, including combinations of causes that are jointly sufficient, a large number of cases usually is not needed. One or two cases may be enough for the simple purpose of eliminating (though not confirming) an explanation about necessary and sufficient causation. A medium number of cases is normally needed to achieve statistical confidence about the validity of an explanation that invokes necessary and/or sufficient causation solely by using cross-case matching techniques.<sup>6</sup> In some comparative-historical studies, this medium number of cases is analyzed. However, in small-*N* studies (e.g.  $N = 3$ ), cross-case analysis is generally combined with process tracing. Because the *N* needed for necessary and sufficient causation is relatively modest, the “burden” that process tracing must carry in such studies is not overwhelming. Rather, the small-*N* comparison does some of the work, with process tracing contributing the rest.

### 3 IMPLICATIONS OF THE DIFFERENCES

Our discussion has called attention to fundamental differences between comparative-historical analysis and statistical analysis. On the one hand, an awareness of these differences provides a basis for appreciating their distinctive contributions in political science. On the other hand, these differences raise questions about the extent to which the two research traditions might be meaningfully combined. By way of conclusion, we address these implications.

The kinds of knowledge generated by comparative-historical research and statistical research are clearly different. Comparative-historical studies tell us why particular outcomes happened in specific cases—this is one important sense in which these studies are “historical,” though there are others (see Mahoney and Rueschemeyer 2003; Pierson 2004; Skocpol 1984). This historical knowledge, in turn, is relevant for policy and practical reasons. By teaching us about the genesis of outcomes in certain specific cases, the knowledge provides a critical foundation for hypothesizing about the effects of subsequent developments in these cases. Here a comparison with physicians who seek the medical history of their patients is useful. A cardiologist can offer better advice to a patient if the causes of the patient’s earlier heart attack are well understood. Analogously, policy-makers can pursue better interventions and

combinations of causes may each be sufficient, such that there are multiple causal paths to the same outcome (see Ragin 1987).

<sup>6</sup> Using Bayesian assumptions, for example, Dion (1998) shows that only five cases may be enough to yield 95% confidence about necessary causes. Using a simple binomial probability test, Ragin (2000, 113–15) shows that if one works with “usually necessary” or “usually sufficient” causes, seven consistent cases are enough to meet this level of significance. Braumoeller and Goertz (2000) offer many examples of case-oriented studies that pass such significance tests.

offer more helpful suggestions if they understand well the causes of prior relevant outcomes in the cases of interest. Indeed, if one understands a particular pattern of causation in a given case, one would seem especially well situated to explore whether the causal pattern might apply to another similar case. These points will be obvious to some, but the tendency for many in the discipline is nevertheless to assume that comparative-historical studies are of mostly historical relevance alone.

The strengths and payoffs of statistical research are different. Whereas comparative-historical analysis is excellent at engaging complex theories with fine-grained over-time evidence, statistical research has the virtue of allowing for the testing of hypotheses about the average effects of particular variables (or specified interactions of variables) within large populations in a way that mimics aspects of a controlled experiment. Findings from large populations may or may not be relevant for thinking about particular cases. For example, a causal variable that promotes a given outcome in the population as a whole might have the opposite effect in a particular case of interest. But statistical findings certainly are relevant for generalizing. Indeed, if one wishes to offer policy advice or recommendations that are intended to—on average—make changes across a large population, the findings generated from statistical methods would seem especially appropriate.

This discussion is not intended to suggest that statistical work is irrelevant for thinking about particular cases. Nor is it meant to suggest that comparative-historical works cannot arrive at quite general findings. Rather, the point is that comparative-historical and statistical studies have different goals, produce different kinds of information, and thus *tend to be* useful for different (though equally valid) purposes.

Given that each tradition has its own distinctive contributions to make, it is not surprising that there would be interest in combining the two, which perhaps could allow for a “best of both worlds” synthesis. While contemporary political scientists often value multimethod research, we nevertheless wish to raise here some cautionary notes about combining comparative-historical analysis and statistical analysis. We believe that the combination is more difficult to achieve than is sometimes suggested, and that multimethod research is not always an improvement over work that is exclusively comparative-historical or exclusively statistical.

When they engage in multimethod research, most analysts still pursue either a causes-of-effects approach or an effects-of-causes approach. In this sense, much multimethod research can be considered *primarily* comparative-historical or *primarily* statistical in orientation. With multimethod work that is primarily comparative-historical, the main goal remains the explanation of specific outcomes in particular cases. The statistical analysis is subservient to this goal. By contrast, with multimethod work that is primarily statistical, the main goal is to estimate average causal effects for a large population. Here one or more case studies are used to service this larger goal. Occasionally, of course, some studies will pursue both goals equally and thus truly cross the divide. However, in our sample, this kind of multimethod research characterized only 8.7 percent of all journal articles.

How is statistical analysis used in multimethod studies that are primarily comparative-historical in orientation? In the most basic way, generalizations from

prior statistical research represent background knowledge that comparative-historical analysts must consider as they formulate their own explanatory hypotheses for their case studies. All comparative-historical analysts react to prior general theories relevant to their outcomes, which often entails situating one's argument in relationship to existing statistical knowledge. Beyond this, comparative-historical researchers also may use statistical findings—including findings they generate themselves—in conjunction with process tracing. Much as a detective draws on knowledge of general causal principles to establish a link between suspect and crime, so too a comparative-historical researcher may use existing or newly discovered statistical findings when attempting to establish the mechanisms that connect cause and effect. For example, one might hypothesize that slow increases in grain prices in eighteenth-century France contributed to peasant revolts by deflating rural wages (i.e. the impact of declining grain prices on overall revolts worked through lower wages at the individual level). To develop this idea, a comparative-historical researcher might wish to carry out regression analysis to assess the effects of prices on wages in France—to make sure that the two are, in fact, statistically linked net of other factors (see Goldstone 1991, 188–9). In doing this, the researcher collects a large number of observations from what is, given the perspective of the comparative-historical research design, a single case. Comparative-historical researchers thus may be especially likely to turn to statistical analysis when macro-hypotheses in the small-*N* research design suggest mechanisms that work at lower levels of analysis. The statistical confirmation of these hypotheses serves the larger goal of validating the small-*N* argument.

For their part, statistical researchers may draw on the findings from comparative-historical analysis to develop their own hypotheses; comparative-historical work can inspire new ideas about causally relevant factors that can be tested in a statistical model. Statistical researchers may also turn to case studies to determine whether findings make sense when assessed in light of an intensive analysis of specific cases. Through such analyses, statistical researchers can evaluate whether the statistical model is adequate, needs refining and retesting, or is deeply problematic and cannot be salvaged. Although in the course of the case analyses the researcher could potentially seek to develop fully adequate explanations of the particular cases, the overarching goal typically remains estimating the average effects of independent variables of interest for the population as a whole. For instance, in Lieberman's (2005) nested analysis approach, cases are selected not because their outcomes are inherently interesting, but rather because their location with respect to the regression line makes them good candidates for further assessing the validity of the statistical model. The goal of the nested analysis is generating valid knowledge about effects of causes; the comparative-historical evidence is mostly subordinated to the larger statistical design.

Our purpose in noting that one approach typically is subordinated to the other in multimethod research is not intended as a criticism. Rather, we emphasize the point to make it clear that most multimethod research is not equal parts quantitative and qualitative—it is, rather, driven by primarily the goals and orientations of one side or the other. When this point is acknowledged, it becomes clear that multimethod

research is an advantage only to the extent that the use of the secondary method actually and effectively supplements the main method of investigation. Statistical studies that offer superficial case studies as supporting evidence do not contribute to the explanation of particular outcomes in those cases. And if the case studies are carried out without attention to good methodological practice, they will not provide a reliable basis for evaluating the statistical model either. By the same token, comparative-historical studies that use regression analysis in the course of process tracing are not necessarily more powerful than comparative-historical studies that do not use any statistical testing. The value added by statistical testing simply depends on what kind of evidence is needed for successful process tracing to be carried out. And the use of regression analysis with process tracing will not be fruitful if the regression analysis is poorly executed.

The message of this discussion is that there is nothing inherently wrong with conducting comparative-historical work that does not include a statistical component (and vice versa). Indeed, for many research projects, an additional secondary analysis using an alternative methodology is unnecessary or inappropriate. Hence, as political science increasingly moves toward and celebrates multimethod research, we believe that some of the best work that is produced in the discipline will eschew this trend and remain squarely centered in the field of comparative-historical analysis.

## APPENDIX A: A NOTE ON CODING PROCEDURES

.....

All nine attributes are measured dichotomously as present or absent in a given article. We allow that, in principle, any study could possess any combination of attributes. Brief operational definitions for the nine attributes follow.

- (1) *Causes-of-effects approach*: present if a central goal of analysis is to explain one or more specific outcomes in particular cases.
- (2) *Effects-of-causes approach*: present if a central goal of analysis is to estimate the extent to which particular independent variables account for variation in the dependent variable for a population as a whole rather than any particular case.
- (3) *Necessary and sufficient conception of causation*: present if individual causal factors or sets of multiple causal factors are treated as necessary and/or sufficient for the outcome of interest.
- (4) *Average effects conception of causation*: present if individual causes are assumed to exert average symmetrical effects that operate within the population as a whole.
- (5) *Process-tracing method*: present if specific pieces of data are used to test the mechanisms through which a potential causal factor is hypothesized to contribute to the outcome of interest.
- (6) *Regression method*: present if regression techniques are used to test hypotheses.
- (7) *Comparative set-theoretic methods*: present if two or more cases are compared across causal and outcome variables to assess whether potential causal variables can be *logically* eliminated.

- (8) *Temporal sequencing or path dependence*: present if the timing or sequencing of independent variables is hypothesized to affect the outcome, or if path-dependent processes are assumed to be present.
- (9) *Rational choice framework*: present if theories are either formally or informally deduced from the assumption of rational, goal-oriented actors.

## REFERENCES

- ACHEN, C. H., and SNIDAL, D. 1989. Rational deterrence theory and comparative case studies. *World Politics*, 41: 143–69.
- BENDIX, R. 1974. *Work and Authority in Industry: Ideologies of Management in the Course of Industrialization*. Berkeley: University of California Press.
- BENNETT, A. 2006. Stirring the frequentist pot with a dash of Bayes. *Political Analysis*, 14: 339–44.
- BENSEL, R. F. 1990. *Yankee Leviathan: The Origins of Central State Authority in America, 1859–1877*. New York: Cambridge University Press.
- BRADY, H. E., and COLLIER, D. (eds.) 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Md.: Rowman and Littlefield.
- BRAUMOELLER, B. F. 2003. Causal complexity and the study of politics. *Political Analysis*, 11: 209–33.
- 2004. Hypothesis testing and multiplicative interaction terms. *International Organization*, 58: 807–20.
- and GOERTZ, G. 2000. The methodology of necessary conditions. *American Journal of Political Science*, 44: 844–58.
- BUNCE, V. 1999. *Subversive Institutions: The Design and the Destruction of Socialism and the State*. Cambridge: Cambridge University Press.
- CLARK, W. R., GILLIGAN, M. J., and GOLDER, M. 2006. A simple multivariate test for asymmetric hypotheses. *Political Analysis*, 14: 311–31.
- COLLIER, D. 1998. Comparative-historical analysis: where do we stand? *APSA-CP: Newsletter of the Organized Section in Comparative Politics*, 9: 1–2, 4–5.
- BRADY, H. E., and SEAWRIGHT, J. 2004. Sources of leverage in causal inference: toward an alternative view of methodology. Pp. 229–66 in Brady and Collier 2004.
- and MAHONEY, J. 1996. Insights and pitfalls: selection bias in qualitative research. *World Politics*, 49: 56–91.
- — and SEAWRIGHT, J. 2004. Claiming too much: warnings about selection bias. Pp. 85–102 in Brady and Collier 2004.
- COLLIER, R. B. 1999. *Paths toward Democracy*. New York: Cambridge University Press.
- and COLLIER, D. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton, NJ: Princeton University Press.
- COPPEDGE, M. 2007. *Approaching Democracy: Research Methods in Comparative Politics*. New York: Cambridge University Press.
- DION, D. 1998. Evidence and inference in comparative case study. *Comparative Politics*, 30: 127–46.
- DOWNING, B. M. 1992. *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. Princeton, NJ: Princeton University Press.
- EKIERT, G. 1996. *The State against Society: Political Crises and their Aftermath in East Central Europe*. Princeton, NJ: Princeton University Press.
- ELMAN, C. 2005. Explanatory typologies in qualitative studies of international politics. *International Organization*, 59: 293–326.
- ERTMAN, T. 1997. *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. Cambridge: Cambridge University Press.
- ESPING-ANDERSON, G. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.
- EVANS, P. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton, NJ: Princeton University Press.
- FREEDMAN, D. A. 1991. Statistical models and shoe leather. In *Sociological Methodology*, ed. P. Marsden. San Francisco: Jossey-Bass.
- GEDDES, B. 1990. How the cases you choose affect the answers you get: selection bias in comparative politics. Pp. 131–50 in *Political Analysis*, vol. ii, ed. J. A. Stimson. Ann Arbor: University of Michigan Press.
- 1991. Paradigms and sand castles in comparative politics of developing areas. Pp. 45–75 in *Comparative Politics, Policy, and International Relations*, ed. W. Crotty. Evanston, Ill.: Northwestern University Press.
- 2003. *Paradigms and Sand Castles: Theory Building in Comparative Politics*. Ann Arbor: University of Michigan Press.
- GEORGE, A. L., and BENNETT, A. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass.: MIT Press.
- GOERTZ, G., and MAHONEY, J. 2007. Scope in case study research. Manuscript.
- and STARR, H. (eds.) 2003. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, Md.: Rowman and Littlefield.
- GOLDSTONE, J. A. 1991. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press.
- 1997. Methodological issues in comparative macrosociology. *Comparative Social Research*, 16: 107–20.
- GOLDTHORPE, J. H. 1997. Current issues in comparative macrosociology: a debate on methodological issues. *Comparative Social Research*, 16: 1–26.
- GOODWIN, J. 2001. *No Other Way Out: States and Revolutionary Movements, 1945–1991*. Cambridge: Cambridge University Press.
- HAGGARD, S. 1990. *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries*. Princeton, NJ: Princeton University Press.
- HICKS, A. 1999. *Social Democracy and Welfare Capitalism: A Century of Income Security Politics*. Ithaca, NY: Cornell University Press.
- HUBER, E., and STEPHENS, J. D. 2001. *Development and Crisis of the Welfare State: Parties and Policies in Global Markets*. Chicago: University of Chicago Press.
- KARL, T. L. 1997. *The Paradox of Plenty: Oil Booms and Petro-States*. Berkeley: University of California Press.
- KING, G., KEOHANE, R. O. and VERBA, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- KOHLI, A. 2004. *State-directed Development: Political Power and Industrialization in the Global Periphery*. New York: Cambridge University Press.
- LIEBERMAN, E. S. 2005. Nested analysis as a mixed method strategy for comparative research. *American Political Science Review*, 99: 435–52.
- LIEBERSON, S. 1985. *Making it Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.
- 1991. Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases. *Social Forces*, 70: 307–20.

- LIEBERSON, S. 1994. More on the uneasy case for using Mill-type methods in small-N comparative studies. *Social Forces*, 72: 1225–37.
- 1998. Causal analysis and comparative research: what can we learn from studies based on a small number of cases. Pp. 129–45 in *Rational Choice Theory and Large-Scale Data Analysis*, ed. H.-P. Blossfeld and G. Prein. Boulder, Colo.: Westview.
- LINZ, J. J., and STEPAN, A. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore: Johns Hopkins University Press.
- LIPJHART, A. 1971. Comparative politics and the comparative method. *American Political Science Review*, 65: 682–93.
- LIPSET, S. M., and ROKKAN, S. (eds.) 1968. *Party Systems and Voter Alignments: Cross-national Perspectives*. New York: Free Press.
- LUEBBERT, G. M. 1991. *Liberalism, Fascism, or Social Democracy: Social Classes and the Political Origins of Regimes in Interwar Europe*. New York: Oxford University Press.
- LUSTICK, I. 1993. *Unsettled States, Disputed Lands: Britain and Ireland, France and Algeria, Israel and the West Bank–Gaza*. Ithaca, NY: Cornell University Press.
- MCKEOWN, T. J. 1999. Case studies and the statistical worldview. *International Organization*, 53: 161–90.
- MACKIE, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- MAHONEY, J. 2000. Strategies of causal inference in small-N analysis. *Sociological Methods and Research*, 28: 387–424.
- 2001. *The Legacies of Liberalism: Path Dependence and Political Regimes in Central America*. Baltimore: Johns Hopkins University Press.
- 2007. Debating the state of comparative politics: views from qualitative research. *Comparative Political Studies*, 40: 32–8.
- and GOERTZ, G. 2006. A tale of two cultures: contrasting qualitative and quantitative research. *Political Analysis*, 14: 227–49.
- and RUESCHEMEYER, D. 2003. Comparative historical analysis: achievements and agendas. Pp. 3–38 in *Comparative Historical Analysis in the Social Sciences*, ed. J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.
- and VILLEGAS, C. 2007. Historical enquiry and comparative politics. Pp. 73–89 in *The Oxford Handbook of Comparative Politics*, ed. C. Boix and S. C. Stokes. Oxford: Oxford University Press.
- MARX, A. W. 1998. *Making Race and Nation: A Comparison of South Africa, the United States, and Brazil*. Cambridge: Cambridge University Press.
- MOORE, B., JR. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.
- PIERSON, P. 1994. *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge: Cambridge University Press.
- 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton, NJ: Princeton University Press.
- RAGIN, C. C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- 2004. Turning the tables: how case-oriented research challenges variable-oriented research. Pp. 123–38 in Brady and Collier 2004.
- RUESCHEMEYER, D., STEPHENS, E. H., and STEPHENS, J. D. 1992. *Capitalist Development and Democracy*. Chicago: University of Chicago Press.

- SEAWRIGHT, J., and COLLIER, D. 2004. Glossary. Pp. 273–313 in Brady and Collier 2004.
- SIKKINK, K. 1991. *Ideas and Institutions: Developmentalism in Brazil and Argentina*. Ithaca, NY: Cornell University Press.
- SKOCPOL, T. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- 1984. Sociology's historical imagination. Pp. 1–21 in *Vision and Method in Historical Sociology*, ed. T. Skocpol. Cambridge: Cambridge University Press.
- 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- 2003. Doubly engaged social science: the promise of comparative historical analysis. Pp. 407–28 in *Comparative-historical Analysis in the Social Sciences*, ed. J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.
- STEINMO, S. 1993. *Taxation and Democracy: Swedish, British and American Approaches to Financing the Modern State*. New Haven, Conn.: Yale University Press.
- TILLY, C. (ed.) 1975. *The Formation of National States in Western Europe*. Princeton, NJ: Princeton University Press.
- 1990. *Coercion, Capital, and European States, AD 990–1990*. Cambridge, Mass.: Basil Blackwell.
- WALDNER, D. 1999. *State-building and Late Development*. Ithaca, NY: Cornell University Press.
- WICKHAM-CROWLEY, T. 1992. *Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes since 1956*. Princeton, NJ: Princeton University Press.
- YASHAR, D. J. 2005. *Contesting Citizenship in Latin America: The Rise of Indigeneous Movements and the Postliberal Challenge*. New York: Cambridge University Press.