A tale of two cultures: contrasting quantitative and qualitative research

James Mahoney
Departments of Political Science and Sociology
Northwestern University
Evanston, IL 60208-1006
email: James-Mahoney@northwestern.edu
and

Gary Goertz
Department of Political Science
University of Arizona
Tucson, Arizona 85721
email: ggoertz@u.arizona.edu

January 27, 2006

Abstract

The quantitative and qualitative research traditions can be thought of as distinct cultures marked by different practices, beliefs, and norms. In this essay, we adopt this imagery toward the end of contrasting these research traditions across ten areas: (1) approaches to explanation, (2) conceptions of causation, (3) multivariate explanations, (4) equifinality, (5) scope and causal generalization, (6) case selection, (7) weighting observations, (8) substantively important cases, (9) lack of fit, and (10) concepts and measurement. We suggest that an appreciation of the alternative assumptions and goals of the traditions can help scholars avoid misunderstandings and contribute to more productive "cross-cultural" communication in political science.

Introduction

Comparisons of the quantitative and qualitative research traditions sometimes call to mind religious metaphors. In his commentary for this issue, for example, Beck likens the traditions to the worship of alternative gods. Schrodt (this issue), inspired by Brady's (2004a, 53) prior casting of the controversy in terms of theology versus homiletics, is more explicit: "while this debate is not in any sense about religion, its dynamics are best understood *as though* it were about religion. We've always known that, it just needed to be said."

We prefer to think of the two traditions as alternative cultures. Each has its own values, beliefs, and norms. Each is sometimes privately suspicious or skeptical of the other though usually more publicly polite. Communication across traditions tends to be difficult and marked by misunderstanding. When members of one tradition offer their insights to members of the other community, the advice is likely to be viewed (rightly or wrongly) as unhelpful and even belittling.

As evidence, consider the reception of Ragin's *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (1987) and King, Keohane, and Verba's *Designing Social Inquiry: Scientific Inference in Qualitative Research* (1994). Although Ragin's book was intended to combine qualitative and quantitative methods, it was written from the perspective of a qualitative researcher, and it became a classic in the field of qualitative methodology. However, statistical methodologists largely ignored Ragin's ideas, and when they did engage them, their tone was often quite dismissive (e.g., Lieberson 1991, 1994; Goldthorpe 1997). For its part, King, Keohane, and Verba's famous work was explicitly about qualitative research, but it assumed that quantitative researchers have the best tools for making scientific inferences, and hence qualitative researchers should attempt to emulate these tools to the degree possible. Qualitative methodologists certainly did not ignore King, Keohane, and Verba's work. Instead, they reacted by scrutinizing the book in great detail, pouring over each of its claims, and sharply criticizing many of its conclusions (e.g., see the essays in Brady and Collier 2004).

Table 1: Contrasting qualitative and quantitative research

Section	Criterion	Qualitative	Quantitative
1	Approaches to explanation	explain individual cases; "causes-of-effects" approach	estimate average effect of independent variables; "effects-of-causes" approach
2	Conceptions of causation	necessary and sufficient causes; mathematical logic	correlational causes; probability/statistical theory
3	Multivariate explanations	INUS causation; occasional individual effects	additive causation; occasional interaction terms
4	Equifinality	core concept; few causal paths	absent concept; implicitly large number of causal paths
5	Scope and generalization	adopt a narrow scope to avoid causal heterogeneity	adopt a broad scope to maximize statistical leverage and generalization
6	Case selection practices	oriented toward positive cases on dependent variable; no (0,0,0) cases	random selection (ideally) on independent variables; all cases analyzed
7	Weighting observations	theory evaluation sensitive to individual observations; one misfit can have important impact	all observations are a priori equally important; overall pattern of fit is crucial
8	Substantively important cases	substantively important cases must be explained	substantively important cases not given special attention
9	Lack of fit	non-conforming cases are examined closely and explained	non-systematic causal factors are treated as error
10	Concepts and measurement	concepts center of attention; error leads to concept revision	measurement and indicators center of attention; error is modeled and/or new indicators identified

In this essay, we tell a tale of these two cultures. We do so from the perspective of qualitative researchers who seek to communicate with quantitative researchers.[1] Our goal is to contrast the assumptions and practices of the two traditions toward the end of enhancing cross-tradition communication. Like Brady and Collier (2004), we believe that qualitative and quantitative scholars share overarching goals concerning the desire to produce valid descriptive and causal inferences. Yet, we also believe that these scholars pursue different specific research goals, which in turn produce different norms about research practices. Hence, we emphasize here to a greater degree than Brady and Collier the distinctiveness in basic goals and practices in the two traditions. Having said this, however, we wish to stress that our intention is not to criticize either quantitative or qualitative researchers. In fact, we argue throughout that the dominant practices of both traditions make good sense given their respective goals.[2]

We adopt a criterial approach (Gerring 2001) to thinking about differences between the two traditions and contrast them across ten areas: (1) approaches to explanation, (2) conceptions of causation, (3) multivariate explanations, (4) equifinality, (5) scope and causal generalization, (6) case selection practices, (7) weighting observations, (8) substantively important cases, (9) lack of fit, and (10) concepts and measurement. There are certainly other differences across the two traditions,[3] but our experience has been that these areas are especially important in generating misunderstandings and miscommunication. Table 1 provides a guide to the discussion that follows.

Before proceeding, we should note that our discussion presents a stylized view of both qualitative and quantitative research. Our characterizations are intended to describe dominant norms and practices. One can easily find examples of research in one tradition in which the analyst carries out practices that characterize the other tradition. However, we suggest that most researchers in political science will locate themselves predominantly in one column of table 1. Of course, as with all cultures there will some individuals that have fairly strong attachments to both. And of course there are other research cultures in political science that are not considered here, including especially other qualitative orientations, such as works of interpretive analysis, many ethnographies, and case studies that do not pursue causal inference as a leading research goal.

1 Approaches to explanation

A core goal of qualitative research is the explanation of outcomes in individual cases. For example, qualitative researchers attempt to identify the causes of exceptional

[1] One of the authors of this paper is a quantitative researcher who also writes about qualitative methods; the other is a qualitative researcher and methodologist.

[2] Although we do not pursue it here, we believe that the two traditions often perform well on different criteria associated with good social science methodology (see especially Gerring's (2001) excellent discussion of these criteria).

[3] Some other potential differences concern level of measurement, type of probabilistic approach, understandings of time, importance of path dependence, and rationales for being concerned about omitted variables.

growth in South Korea and Taiwan, the onset of World War I, the end of the Cold War, the creation of generous welfare states in Scandinavia, and the rise of neopopulism in Peru and Venezuela. A central purpose of research is to identify the causes of these outcomes in the particular cases of interest.

By starting with cases and their outcomes and then moving backward toward the causes, qualitative analysts adopt a "causes-of-effects" approach to explanation. Good theories must ideally explain the outcome in all of the cases under investigation. For instance, Skocpol's (1979) theory of social revolution is intended to adequately explain France, Russia, and China; the assessment of the theory, in turn, is based primarily on how well it succeeds at this research objective.

From the qualitative perspective, this approach to asking and answering questions corresponds to ordinary language understandings of causation and is consistent with normal science as conventionally understood. For example, researchers in the fields of evolutionary biology and astronomy often seek to identify the causes of particular outcomes. Indeed, most natural scientists would find it odd that their theories cannot be used to explain individual events. Questions such as "Why did the space shuttle Challenger explode?" are a request for a cause of an effect. When testifying in front of Congress, Richard Feymann did not think this question to be nonsensical or nonscientific (Vaughan 1986).

In contrast, statistical approaches to explanation usually use the paradigm of the controlled experiment.[4] With a controlled experiment, one does not know the outcome until the treatment has been applied. Indeed, one might say that the whole point of the experiment is to observe the effect (if any) of the treatment.

Statistical approaches attempt to reproduce the paradigm of the controlled experiment in the context of an observational study. Although there are important and well-known difficulties in moving from controlled experiment to observational study (e.g., the absence of true randomization and manipulation), for our purposes the crucial point is that statistical researchers follow the "effects-of-causes" approach employed in experimental research. In particular, with a statistical research design, one seeks to estimate the average effect of one or more causes across a population of cases. The explanation of *specific outcomes in particular cases* is not a central concern. Hence, quantitative researchers formulate questions such as "What is the effect of economic development on democracy?" or "What effect does a given increase in foreign direct investment have on economic growth?" They do not normally ask questions such as, "Was economic development a necessary cause of democracy in Chile?" or "Were high levels of foreign investment in combination with soft authoritarianism and export-oriented policies sufficient for the economic miracles in South Korea and Taiwan?"

Methodologists working in the statistical tradition have seen clearly the difference between the causes-of-effects approach, in which the research goal is to explain outcomes in particular cases, and the effects-of-causes approach, in which the research goal is to estimate average effects across a population. In general, however,

[4]For example, Angrist, Imbens, and Rubin (1996, 144) assert that, "causal inference in statistics, going back at least to work by Fisher (1981, 1925) and Neyman (1923) on agricultural experiments, is fundamentally based on the randomized experiment (see also Kempthorne 1952 and Cox 1958)."

they have expressed skepticism about the causes-of-effects approach. For example, Holland responded to comments on his article as follows:

> I must disagree with Gylmour's paraphrasing of my (i.e., Rubin's) analysis, however, and with the counterfactual analysis of causation of Lewis described by Gylmour. I believe that there is an unbridgeable gulf between Rubin's model and Lewis's analysis. Both wish to give meaning to the phrase "*A* causes *B*." Lewis does this by interpreting "*A* causes *B*" as "*A* is *a* cause of *B*." Rubin's model interprets "*A* causes *B*" as "the *effect* of *A* is *B*." (Holland 1986b, 970)

King, Keohane, and Verba (1994) follow Holland quite closely, and they explicitly define causality in terms of the effects-of-causes approach.[5] They do not consider or discuss the causes-of-effects approach to explanation.

The distinction between causes-of-effects and effects-of-causes arises several times in the symposium on Brady and Collier (2004) in this special issue. For example, Beck in his contribution believes it is essential to be clear "whether our interest is in finding some general lawlike statements or in explaining a particular event." In the case of Stokes's (2001) and Brady's (2004b) work, he concedes that "the qualitative analysis is helpful for understanding one specific case," but his basic view advocates looking for effects across large populations. Likewise, Shivley (this issue) suggests that scholars who work with a small number of cases "devote themselves to process-tracing, not to quasi-statistical generalization." His view of causation too is from the effects-of-causes tradition.

We believe that much misunderstanding between the two traditions derives from these different approaches to explanation. Quantitative researchers may have difficulty appreciating the concern of qualitative researchers with explaining outcomes in particular cases. For example, the idea that Skocpol (1979) would really want to write a whole book that is primarily an effort to explain the occurrence of social revolution in only France, Russia, and China may seem puzzling within the statistical culture. "Real science should seek to generalize about causal effects," might be a typical reaction. Yet, from the qualitative perspective, science can precisely be used in service of explaining outcomes in particular cases.

We believe that both approaches are of value; in fact, they compliment one another. Ideally, an explanation of an outcome in one or a small number of cases leads one to wonder if the same factors are at work in other cases, stimulating a larger N analysis in which the goal is less to explain particular cases and more to estimate average effects. Likewise, when statistical results about the effects of causes are reported, it seems natural to ask if these results make sense in terms of the history of individual cases; one wishes to try to locate the effects in specific cases. This complimentarity is what makes possible mixed-method research in which scholars combine

[5]In contrast to Holland, Dawid (2000) does not go so far as to reject the causes-of-effects approach. Instead, he treats it as a special case of causation. Interestingly, in his response to a series of comments from several distinguished statisticians, he expresses surprise that his analysis of causes-of-effects provoked so little discussion since he thought it would be controversial. "I am surprised at how little of the discussion relates to my suggestions for inference about 'causes of effects,' which I expected to be the most controversial" (Dawid 2000, 446).

large-N and small-N research designs (for recent discussions of mixed-method research strategies, see George and Bennett 2005; Coppedge forthcoming; Lieberman 2005).

2 Conceptions of causation

In order to explain outcomes in particular cases, qualitative researchers often think about causation in terms of necessary and/or sufficient causes. The adoption of this understanding of causation can be seen clearly in the kinds of comparative methods employed by qualitative researchers. Mill's methods of difference and agreement, explanatory typologies, and Ragin's qualitative comparative methods are all predicated in one way or another on necessary and/or sufficient causation (see George and Bennett 2005; Goertz and Starr 2003; Ragin 1987, 2000).[6]

From the qualitative perspective, the assessment of necessary and/or sufficient causation seems quite natural and fully consistent with logic and good science. For example, classical qualitative methodologists – such as Weber (1949), Aron (1986), and Honoré and Hart (1985), in fact going back to David Hume – think about causation in individual cases in terms of a necessary condition counterfactual: if $\neg X$ then $\neg Y$. X is a cause of Y because without X, Y would not have occurred. This approach to causation corresponds to the preference of most qualitative analysts for expressing their theories using logic and set-theoretic terms. Likewise, as various methodologists point out, this understanding of causation is common in historical explanation:

> If some event A is argued to have been the cause of a particular historical event B, there seems to be no alternative but to imply that a counterfactual claim is true – if A had not occurred, the event B would not have occurred.[7] (Fearon 1996, 40; see also Nagel 1961, 581–82; Gallie 1955, 161)

When qualitative scholars move from individual cases to a small or medium N, they often adopt the "INUS" approach to causation (Mackie 1980; Ragin 1987, 2000).[8] An INUS cause is neither individually necessary nor individually sufficient for an outcome. Instead, it is one cause within a combination of causes that are jointly sufficient for an outcome. Thus, with this approach, scholars seek to identify combinations of variable values that are sufficient for outcomes of interest. The approach assumes that distinct combinations may each be sufficient, such that there are multiple causal paths to the same outcome (this is sometimes called equifinality; see below). In this tradition, comparative qualitative research findings can often be formally expressed through Boolean equations such as Y=(A AND B AND C) OR (C AND D AND E).

The situation is quite different on the quantitative, statistical side. Here the analyst typically seeks to identify causes that, on average, affect (e.g., increase or

[6]As qualitative researchers increase their N and start to measure variables on an ordinal scale, however, they may shift to a correlational understanding of causation (see Mahoney 2000).

[7]Note that the problem of historical explanation is expressed in causes-of-effects terms.

[8]An INUS condition is a "an *insufficient* but *nonredundant* part of an *unnecessary* but *sufficient* [combination of conditions]" (Mackie 1980, 62).

decrease) the values on an outcome across a large population. For convenience, we call this the correlational approach to causation. More formally, one can define this approach to causation for a single case in terms of a counterfactual: the difference between the treatment (T) and control (C) for the same unit, i. Using the framework and notation of King, Keohane, and Verba (1994), we have for an individual case:

$$\text{Causal Effect} = y_i^T - y_i^C \quad \text{T} - \text{treatment, C} - \text{control} \tag{1}$$

This equation represents what King, Keohane, and Verba (1994, 78–79) call the "realized causal effect" for unit i (Dawid 2000 calls this the "individual causal effect"). Unlike the logical and set-theoretic focus of qualitative research, the quantitative approach uses an *additive* criterion to define cause: $y_i^T - y_i^C$.

When the quantitative approach moves from the individual case to multiple cases, the understanding of causal effect as an (unobservable) contrast between control and treatment for an individual observation becomes the causal effect for multiple observations through the comparison of *groups*, in other words over many units $i = 1, \ldots, N$. Again using the basic notation of King, Keohane, and Verba:

$$\text{Mean Causal Effect} = \mu^T - \mu^C \quad \text{T} - \text{treatment, C} - \text{control} \tag{2}$$

Not surprisingly, King, Keohane, and Verba refer to this as β.[9] This is variously called the "mean causal effect" (Holland 1986a), "average treatment effect" (Sobel 2000), "average causal response" (Angrist and Imbens 1995), or "average causal effect" (Dawid 2000). Thus, the statistical approach replaces the impossible-to-observe causal effect of T on a specific unit with the possible-to-estimate *average* causal effect of T over a population of units (Holland 1986a, 947). Hence it is an easy step to consider causal effects as being the βs one estimates in statistical models.

Given these different conceptualizations of causation, there is real potential for misunderstanding and miscommunication. In fact, the kinds of hypotheses developed in the two traditions are not always commensurate. For example, consider Waldner's (1999) hypotheses about state building and economic development in Turkey, Syria, Taiwan, and Korea: low levels of elite conflict and a narrow state coalition are both necessary for a developmental state; a developmental state in turn is necessary and sufficient for sustained high growth. It is not clear how a scholar working within statistical framework would evaluate or understand these causal claims. Possibly, she would translate the hypotheses into language that is familiar to her. Thus, she might assume that Waldner hypothesizes that: (1) elite conflict and coalitional shallowness are positively associated with the presence of a developmental state, and (2) a developmental state is positively associated with economic development. But Waldner does not in fact develop (or necessarily agree with) these hypotheses; his argument cannot be unproblematically translated into the language of correlational causation.

The reaction of statistical researchers to the qualitative approach to causation is often one of profound skepticism. This skepticism may be grounded in the belief that there are no necessary and/or sufficient causes of social phenomena, that these kinds

[9]Actually King, Keohane, and Verba (1994) use β to refer to the mean causal effect for unit i, which we would notate as β_i.

of causes make untenable deterministic assumptions, or that these kinds of causes must be measured as dichotomies.[10] Statistical researchers may therefore choose to dismiss out of hand qualitative hypotheses that assume necessary/sufficient causation. Alternatively, as suggested with the Waldner example, they may choose to reinterpret them as representing implicit correlational hypotheses.

Our view is that it is a mistake to reject in toto alternative understandings and definitions of cause. For one thing, there are in fact different mathematical models for representing the idea of cause within each tradition. For example, within the statistical tradition, one does not have to define causal effects in additive terms. Rather, as Dawid (2000) notes, one could use y_i^T/y_i^C or $\log(y_i^T/y_i^C)$. Also, as Braumoeller (this issue) suggests, one could model causal effects as appearing in the variance rather than the mean. Likewise, in the qualitative tradition, one could think of causation in singular cases in terms of sufficiency without necessity: "a [covering, scientific] law has the form IF conditions C1, C2, ...,Cn obtain, THEN always E" (Elster 1999, 5) or "every general proposition of the form 'C causes E' is equivalent to a proposition of the form 'whenever C, then E'" (Ayer 1946, 55). In sum, we argue that there is no a priori reason for dismissing different kinds of causes. Theories regularly posit alternative notions of cause, and scholars should be open to working with different conceptions of causation, depending on the implications of their theory. While to some this may seem self evident, the tendency in political science has too often been to dismiss certain understandings of causation or to use methods that assume an understanding that is not congruent with the theory under investigation (see, e.g., Hall 2000).

3 Multivariate explanations

In all research, the desire to explain leads to a multivariate focus. In qualitative research, this can be seen with the assumption that individual events do not have *a* cause; rather one must include a variety of casually relevant factors. In quantitative research, of course, one normally assumes that it is impossible to estimate average effects without controlling for relevant variables.

Yet the typical multivariate model of each tradition varies in quite important ways. Take perhaps the most common, modal, model in each traditions:

$$Y = (A*B*c) + (A*C*D*E) \quad * = \text{Logical AND} + = \text{Logical OR} \tag{3}$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_{12} X_1 * X_2 + \epsilon \tag{4}$$

Equation (3) represents a typical set-theoretic Boolean model based on the INUS approach to causation (lower-case letters indicate the negation of a variable; with dichotomous variables, this means the absence of the variable). The equation identifies two different combinations of variables that are sufficient for the outcome. By contrast, equation (4) is a standard statistical model that includes an interaction term. Although in some ways these two equations may appear to be similar, there are in fact important differences.

[10] Not surprisingly, qualitative researchers have responded systematically to these kinds of concerns (e.g., Goertz and Starr 2003; Mahoney 2004).

In the qualitative tradition, one often focuses primarily on the impact of combinations of variables and only occasionally focuses on the effects of individual variables. Indeed, unless a variable is a necessary cause or individually sufficient for an outcome, the qualitative researcher will usually make no effort to estimate its net effect. For example, in equation (3) the qualitative researcher would certainly point out that variable A is necessary for the outcome. But it makes virtually no sense to ask, "what is the effect of cause C?" Because C sometimes has a positive effect and sometimes a negative effect depending on the other variable values with which it appears, asking about its "net" effect is not a fruitful approach. Likewise, B matters in the presence of A and c but in other settings it has no effect on the outcome. Hence, it is not useful to generalize about the overall effect of B without saying something about the context (i.e., other variable values) in which B appears.

In the quantitative tradition, by contrast, one is more likely to be focused on estimating the net effect of individual causes, i.e., the individual X_i. For example, in the causal model represented by equation (4), one is centrally concerned with estimating the net effect of each individual variable. To be sure, one can include interaction terms in a statistical model (as we have done). Nevertheless, recent articles on the methodology of statistical interaction terms (Braumoeller 2004; Clarke this issue; Brambor, Clark, and Golder 2006; see also Achen 2005) illustrate that the "individual effect" approach continues to be the norm in statistics as actually practiced in the social sciences. Typically, when scholars use interaction terms they still ask about the individual impact of X (see Braumoeller 2004 for examples and critique).

When a quantitative researcher sees a Boolean model like equation (3), she may instinctively translate it into the familiar terms of interaction effects. This is not a completely unreasonable view (Clark's article in this special issue defends at length this translation), for the logical AND is a first cousin of multiplication. However, a good statistician would almost never actually estimate equation (3). For example, to estimate the model, statistical practice suggests that one should include all lower order terms such as A, AB, AC, and AD. Although there are very good statistical reasons for this practice, in Boolean models these reasons do not exist because one is dealing with logic and set theory.

In fact, the logical AND in equation (3) is not the same as multiplication in equation (4). Nor is the logical OR in equation (3) the same as addition in equation (4). We believe that a failure to recognize these differences contributes to substantial confusion across the two traditions. In particular, it causes quantitative scholars to believe that a Boolean model is a set of interaction terms that could easily be translated into statistical language (e.g., King, Keohane, and Verba 1994, 87–89; Seawright 2005).

One way to illustrate the point is by considering the set-theoretic underpinnings of necessary and sufficient causes. A necessary cause is a superset where the $Y = 1$ cases are one subset (i.e., all cases where the outcome is present are contained within the larger population where the necessary cause is present); a sufficient cause is one subset of a superset of $Y = 1$ cases (i.e., all cases where the sufficient cause is present will be contained within the larger population where the outcome is present).

This set-theoretic logic ensures that there is a consistent relationship at the population level and the subset level for findings that are expressed with the logical AND. For instance, suppose for the whole population that we have a Boolean model such as $Y = (A * b * c) + (A * C)$. Since A is a necessary condition of Y for the whole population, then it must be a necessary condition for any subset of the population.[11] Likewise, if the combination $A * b * c$ is sufficient for the superset, then it must be sufficient for any subset. Of course, $A * b * c$ might not be present in all subsets (e.g., the $A * C$ one). But the point is that if $A * b * c$ is present in a subset, then Y will also be present.

The logical approach of qualitative research can be contrasted with the relationship between populations and subgroups in statistical research. Imagine that in a statistical study the impact of X_1 is strongly positive in the superset (i.e., the full population). Does this mean that X_1 cannot have a strongly negative impact for a particular subset of cases? The answer, of course, is "no." The impact of X_1 as one moves from a superset to subsets is always contingent in statistical models; that is, there is no logically necessary reason why X_1 could not be negatively related to the outcome in particular subsets.[12] Similarly, the estimate of the parameter $\beta_{12} X_1 * X_2$ could change dramatically when moving from the whole population to a subgroup. In short, therefore, there are mathematical differences between the models in equations (3) and (4). A Boolean equation is not the same thing as a statistical interaction model.

The two models represented in equations (3) and (4) are in many ways difficult to compare, which points to real differences across the traditions. But from the perspective of a dialogue between cultures, it is better to understand the differences than to fight over who is right or better. Indeed, the logic and set theory that form the basis of the qualitative view of cause and causal complexity are not more or less rigorous than the probability and statistics used by quantitative scholars.[13] We thus see the two approaches as each viable for social science research.

4 Equifinality

Another clear indicator of differences between the qualitative and quantitative traditions is the importance or lack thereof attributed to the concept of "equifinality" (George and Bennett 2005). Also referred to as "multiple, conjunctural causation"

[11] For a substantive example, take the classic democratic peace hypothesis: democratic dyads do not fight wars. The hypothesis can be phrased in terms of a necessary condition: nondemocracy (i.e., nondemocratic dyads) is a necessary condition for war. Since the set of war dyads is a subset of all nondemocratic dyads, this hypothesis will remain true for any subset of war dyads (i.e., if $A \subset B$ and $B \subset C \Longrightarrow A \subset C$).

[12] The assumptions associated with unit homogeneity and unit independence, e.g., Stable Unit Treatment Value Assumption (see Brady and Seawright 2004 for a nice discussion), are designed to prevent this parameter instability from occurring. In practice, parameter instablility remains a real possibility.

[13] Often the difference is expressed in terms of determinism versus probabilism, but this ignores basic contrasts in functional form that do not concern determinism. Also, many qualitative researchers assume a probabilistic view of causation that is compatible with the set-theoretic focus of their tradition (e.g., Ragin 2000).

(Ragin 1987; Brady and Collier 2004, 278) or just "multiple causation" (King, Keohane, and Verba 1994, 87), the concept of equifinality is strongly associated with the qualitative comparative analysis approach developed by Ragin (1987), and it plays a key role in how many qualitative scholars think about causal relationships. In contrast, discussions of equifinality are absent in quantitative work. If one were to read only large-N quantitative work, the word "equifinality" (or its synonyms) would not be part of one's methodological vocabulary.

Equifinality is the idea that there are multiple causal paths to the same outcome. In terms of multivariate explanations, as we have seen, equifinality is expressed using the INUS approach. In equation (3), for example, there are two causal paths $(A * B * c)$ OR $(A * C * D * E)$; either one is sufficient to attain a positive outcome.

We think that much of the discussion of equifinality inappropriately views its distinctive aspect as the fact that the causal paths are normally represented by combinations of variable values; the causal paths are thus "conjunctural" in nature. If one focuses mainly on this component using a statistical perspective, as do King, Keohane, and Verba (1994), one may (mistakenly, as we have seen) believe that equifinality is simply a way of talking about interactions terms.

What actually makes equifinality distinctive in qualitative work is the fact that *there are only a few causal paths* to a particular outcome. Each path is a specific conjunction of factors, but there are not very many of them. Within the typically more limited scope conditions of qualitative work (see below), the goal is to identify all the causal paths present in the population.

In contrast, implicit in statistical models such as equation (4) are thousands, if not millions, of potential paths to a particular outcome. The right hand side of the statistical equation essentially represents a weighted sum, and as long as that weighted sum is greater than the specified threshold – say in a logit setting – then the outcome should (on average) occur. Within this framework, there will be countless ways that the weighted sum could exceed the threshold. One has equifinality in spades.

In qualitative research, analysts will normally assign cases to causal paths. Since the overall research goal is to explain cases, one does so by identifying the specific causal path that each case follows. For example, Hicks et al. (1995) conclude that there are three separate paths to an early welfare state, and their analysis allows one to identify exactly which cases followed each of the three paths (see also Esping-Andersen 1990). In qualitative research, these causal paths can play a key organizing role for general theoretical knowledge. To cite another example: Moore's (1966) famous work identifies three different paths to the modern world, and the combinations of factors that define these overall paths are essential to understanding the theoretical contribution of this study.[14]

Within quantitative research, it does not seem useful to group cases according to common causal configurations on the independent variables. While one could do this, it is not a practice within the tradition. Again, the goal of research here is

[14]Given that equifinality often organizes causal generalization in qualitative research, it is not surprising that Mackie's (1980) chapter on INUS models is called "causal regularities." With an INUS model, each case may belong to a larger set of cases that follow the same causal path. INUS models thus form a series of theoretical generalizations.

not to explain any particular case, but rather to generalize about causal effects. In this context, one speaks about the population as a whole and does not discuss the particular pathways that individual cases follow to arrive at their specific values on the dependent variable.

5 Scope and causal generalization

In qualitative research, it is common for investigators to define the scope of their theories narrowly such that inferences are generalizable to only a limited range of cases. Indeed, in some qualitative works, the investigator does not seek to generalize findings to cases beyond those analyzed in the study. By contrast, in quantitative research, scholars usually define their scope more broadly and seek to make generalizations about large numbers of cases. Quantitative scholars often view the cases they analyze simply as a sample of a potentially much larger universe.

The narrower scope adopted in qualitative analysis grows out of the conviction that causal heterogeneity is the norm for large populations (e.g., Ragin 1987; 2000). Qualitative researchers assume that as the population size increases, even modestly, the potential for key causal variables to be missing from their theories increases dramatically. Another way of saying the same thing is that these researchers believe that the addition of each case to the analysis stands a good chance of necessitating substantial modifications to the theoretical model, even though the model works perfectly well for the original cases analyzed. Insofar as these modifications produce major complications, qualitative researchers believe that it is better to develop an entirely separate theory for the additional cases. Thus, for example, Skocpol develops separate theories for the great historical social revolutions and for the more contemporary social revolutions in the Third World (Skocpol 1979; Goodwin and Skocpol 1989).

As we saw in the previous section, causal generalization in qualitative methods often takes the form of specifying a few causal paths. Expanding the scope of a study can prove problematic in this context. It might be that the new cases do not fit the current set of causal paths. In terms of equation (3), for example, one has two causal paths ($A*B*c$) OR ($A*C*D*E$), and enlarging the scope might mean that the new cases require the addition of a third or fourth causal path. It can also arise that the new cases make existing causal paths problematic. For example, the path ($A*B*c$) may no longer be sufficient for the outcome of interest in the new scope.

Research practices are quite different in the quantitative tradition. Here of course researchers need to have a large number of observations to use most statistical techniques, which may encourage a broad understanding of theory scope. But more importantly, the very conception of causation used in quantitative research means that the concerns of causal heterogeneity are cast in different terms. In particular, if your goal is to estimate an average effect of some variable or variables, the exclusion of certain variables associated with new cases is not a problem as long as assumptions of conditional independence still hold.[15] Independent variables that are important for only a small subsets of cases may be appropriately considered "unsystematic" and relegated to the error term.[16] Hence, in quantitative research, where adequate explanation does not require getting the explanation right for each case, analysts can omit minor variables to say something more general about the broader populations.

One key implication of this difference is that causal generalizations in qualitative work are much more fragile than those in large N statistical analyses. Statistical analyses are often robust and will not be dramatically influenced by modest changes in scope or population. But in qualitative research, heterogeneity of various sorts (e.g., concepts, measurement, and model) poses a major problem, which in turn makes qualitative scholars particularly likely to restrict the domain of their argument.

These differences should not be used for the purpose of building an indictment against either tradition; rather, they should be seen as natural outgrowths of the traditions' different research goals. In qualitative research, your main goal is to produce good explanations for the particular cases under investigation, and it follows that you are likely to be quite concerned with missing variables and the theoretical complications introduced by new cases. It is thus natural that you are hesitant to test your arguments using substantially larger populations of cases than those already analyzed. By contrast, in quantitative research your main goal is to generalize about the average effects of one or more causes and you worry much less about developing a complete explanation for the population as a whole (though prediction models are still sometimes pursued). It is not surprising in this context that you restrict the scope of your theory more reluctantly and less extensively than in qualitative research.

6 Case selection practices

Qualitative researchers usually start their research by selecting cases where the outcome of interest occurred. This practice is not surprising when we recall that their research goal is the explanation of particular outcomes. If you want to explain certain outcomes in specific cases, it is natural to choose cases that exhibit those outcomes. Although sometimes qualitative researchers may only select these cases, quite commonly they choose both "postitive" and "negative" cases to test theory (see Mahoney and Goertz 2004).

In quantitative research, by contrast, researchers generally select cases without regard for their value on the dependent variable. In fact, for well understood reasons, selecting cases based on their value on the dependent variable can bias findings in statistical research (e.g., Heckman 1976). Quantitative researchers therefore ideally try to choose populations of cases through random selection on independent variables.

[15]Of course, some statistical methodologists do not believe that these assumptions usually hold outside of natural experiments (e.g., Freedman 1991). Yet this concern raises a separate set of issues that are best debated from within the statistical tradition itself.

[16]In this sense, the error term of a typical statistical model may contain a number of variables that qualitative researchers regard as crucial causes in individual cases.

Table 2: Case selection

Y	X_1	X_2
$Y = 1$	1	1
	0	1
	1	0
	0	0
$Y = \mathbf{0}$	1	1
	0	1
	1	0
	0	**0**

These basic differences to case selection have stimulated debate across the two traditions. In the late 1980s and early 1990s, Achen and Snidal (1989) and Geddes (1991) criticized qualitative research designs on the subject of selecting on the dependent variable. This foreshadowed King, Keohane, and Verba's (1994) well-known discussion of the issue, which was especially critical of research designs that lack variance on the dependent variable (i.e., "no variance designs"). By the late 1990s, a number of scholars responded to these criticisms. Regarding no variance designs, methodologists pointed out that if the hypothesis under consideration assumes necessary causation, as is common in qualitative research, the design is appropriate (e.g., Dion 1998; Ragin 2000; Harvey 2003; Braumoeller and Goertz 2000).[17] Likewise, other methodologists (e.g., Collier, Mahoney, and Seawright 2004) insisted that within-case analysis, which relies on causal-process observations (discussed below), provides substantial leverage for causal inference even when the N equals 1. Nevertheless, in many research designs, qualitative scholars include negative outcome cases for the purposes of causal contrast and inference (for example, Skocpol also examines six negative cases where social revolution did not occur in addition to her three positive cases).

To highlight the real differences to case selection in the two traditions, an example is helpful. In table 2, there are two independent variables and one dependent variable; all variables are measured dichotomously. In a standard experimental design, one can manipulate cases such that they assume the four possible combinations of values on the independent variables and then observe their values on the dependent variable. In statistical analysis, the selection of a large number of cases without regard for their value on the dependent variable has the effect of approximating this experimental design.

In the typical small-N study, however, there are two characteristics that are somewhat distinctive. The first is that there are usually very few cases of 1 on

[17] While there is mostly consensus on this point, Braumoeller and Goertz (2000) show that no variance designs do not permit one to distinguish trivial from non-trivial necessary causes. For a different view, see Seawright (2002), who argues for the use of "all cases" and not merely those where $Y = 1$ when testing necessary causation.

the dependent variable; in terms of table 2, the top half of the table is much less populated than the bottom half. This is true because the positive cases of interest (i.e., cases where $Y = 1$) in qualitative research are generally rare occurrences (e.g., wars, revolutions, growth miracles), while the negative cases (e.g., nonwars, nonrevolutions, nongrowth miracles) are potentially almost infinite in size. Of course, the same can be true in experimental or statistical research when analysts study rare events (e.g., see Goldstone et al. 2000; King and Zeng 2001), though as a generalization we can say that the study of exceptional outcomes is more common in qualitative research.

The second and more important distinctive trait of qualitative analysis is that in the heavily populated bottom half, the (0,0,0) cell (in bold type in the table) where both causes and the outcome are absent is particularly heavily populated and problematic. In practice, qualitative researchers rarely choose cases (or case studies) from the (0,0,0) cell. A practical reason why is that the (0,0,0) cases are so numerous and ill-defined that it is difficult to select only a few for intensive analysis, while selecting a large number of these cases is not a realistic option. By contrast, in a statistical analysis, having a lot of cases is desirable, and computation of statistical results is not hindered but helped by having many cases in each cell.

Another problem confronting the qualitative scholar is that the (0,0,0) cases are less useful for testing theories when compared to cases taken from the other cells. For example, assume that the causal model being tested in table 2 is $Y = X_1$ AND X_2. Negative cases in the (0,1,1) cell are extremely useful because they disconfirm or at least count against this theory (i.e., both causes are present, but the outcome is absent); hence, qualitative researchers are highly attuned to finding these cases. Likewise, negative cases in the (0,1,0) and (0,0,1) cells help qualitative researchers illustrate how X_1 and X_2 are not individually sufficient for the outcome. But the (0,0,0) cases provide less leverage for causal inference (Braumoeller and Goertz 2000). In fact, in most of these cases, the outcome of interest is not even possible and thus the cases are regarded as irrelevant (Mahoney and Goertz 2004). In short, one will almost never see a qualitative scholar doing a case study on an observation from the (0,0,0) cell.

In contrast, in quantitative research, increasing variance reduces the standard error and thus is pursued when possible. Within a statistical framework, therefore, one would normally wish to include cases distant from 1 on the independent variables, such as cases from the (0,0,0) cell.

In these important ways, the two traditions differ in how they approach case selection on both the dependent and independent variable sides of the equation. Yet, we are convinced that both traditions have good reasons for doing what they do. If your goal is to estimate average causal effects for large populations of cases, it makes sense to avoid selecting on the dependent variable. Likewise, it makes sense to include all types of negative cases and treat them as equally important for drawing conclusions about causal effects. But if your goal is to explain outcomes in particular cases, it does not make sense to select cases without regard for their value on the outcome. Nor does it make sense to include all negative cases that lack the outcome of interest as equally relevant to the analysis.

7 Weighting observations

Qualitative researchers are in some ways analogous to criminal detectives. Like a detective solving a crime, a qualitative researcher explains a particular outcome by drawing on detailed fact gathering, experience working with similar cases, and knowledge of general causal principles. From the standpoint of this "detective" method (Goldstone 1997; see also McKeown 1999; Van Evera 1997, chap. 2), not all pieces of evidence count equally for building an explanation. Rather, certain observations may be "smoking guns" that contribute substantially to a qualitative researcher's view that a theory is valid. By the same token, much like a detective whose initial hunch about a murder suspect can be undermined by a single new piece of evidence (e.g., an air-tight alibi), qualitative researchers may conclude that a given theory is not correct despite the fact that a considerable amount of evidence suggests that it is. For qualitative researchers, a theory is usually only one critical observation away from being falsified. And yet, one sometimes builds enough evidence to feel quite confident that the theory is valid and that no falsifying evidence will subsequently be found.

Also like detectives, qualitative researchers do not view themselves as approaching observations in a theoretically neutral way. Rather, these researchers in effect ask: "Given my prior theoretical beliefs, how much does this observation affect these beliefs" (Goldstone 2003)? When testing some theories, a single piece of data can radically affect posterior beliefs. The crucial data could show that a key variable was incorrectly measured, and when correctly measured, the theory no longer makes sense. We see this with the theory that held that China performed better than India on key social indicators before 1980 because of its higher level of GDP per capita. When researchers introduced a new measure of economic development, which addressed problems with the previous GDP per capita estimate and showed similar levels of development in the two countries, the whole theory was immediately called into question and rejected (Drèze and Sen 1989). The decisive data need not involve a measurement problem. For instance, consider the theory that the combination of a weak bourgeoisie, a divided peasantry, and a powerful landed elite are sufficient for fascism in interwar Europe (Moore 1966). This theory might be called into question by simply observing that the powerful landed elites in the fascist cases either could not deliver large numbers of votes or were actually supporters of liberal candidates (Luebbert 1991, 308–9). When one recognizes this point, the theory seems deeply problematic, despite the fact it is plausible in other ways (for other examples, see McKeown 1999).

By contrast, quantitative scholars generally make no assumptions that some pieces of evidence – i.e., particular observations – should count more heavily than others. Rather, quantitative researchers usually weight a priori all observations equally. They then work to establish a pattern of conforming observations against a null hypothesis. With this approach, a single observation cannot lend decisive support or critically undermine a theory; only a pattern of many observations can bolster or call into question a theory. Statistical results which draw too heavily on a few specific observations (often outliers) are suspect.

These different uses of data also correspond to Brady and Collier's (2004, 252–55) distinction between "causal-process" and "data-set" observations. A data-set observation is simply a row in a standard rectangular data set and is ordinarily what statistical researchers call a case or an observation. Data-set observations provide analytic leverage because they show or do not show statistically significant patterns of association between variables as well as allow for the estimation of the size of effects. By contrast, "A causal-process observation is an insight or piece of data that provides information about context or mechanism and contributes a different kind of leverage in causal inference. It does not necessarily do so as part of a larger, systematized array of observations.... a causal-process observation may be like a 'smoking gun.' It gives insight into causal mechanisms, insight that is essential to causal assessment and is an indispensable alternative and/or supplement to correlation-based causal inference" (Brady and Collier 2004, 252–53). Causal-process observations are crucial for theory testing in a qualitative setting precisely because one sorts through the data with pre-existing theoretical beliefs (including common sense).

Like Brady and Collier, we believe that both kinds of evidence can be useful. We would simply add that causal-process observations are especially useful when one seeks to explain specific outcomes in particular cases, whereas data-set observations are especially helpful when one wishes to generalize about average causal effects for a large population. Thus, if your goal is to explain particular outcomes, it makes sense to move back and forth between theory and the data; it does not make sense to carry out a single pass of the data or to avoid all ex post model revisions (though researchers must still be sensitive to simply fitting a theory to the data). By contrast, if one seeks to estimate average causal effects, one should normally assume a more strict differentiation between theory and data, and one should not move as freely back and forth between theory and data (though specification searches and other data probes may be consistent with good practice). The upshot is that quantitative researchers should not primarily seek out causal-process observations anymore than qualitative researchers should primarily study data-set observations. Both sets of scholars, rather, should continue what they are doing and work on improving their techniques from within the assumptions of their own tradition.

8 Substantively important cases

Qualitative and quantitative scholars have different perspectives on what constitutes an "important" case. In a typical large-N analysis, there are no *ex ante* important cases. Each case carries equal weight. *Ex post* one can and should examine outliers and observations that have large leverage on the statistical results. And techniques have long existed for identifying and analyzing these kinds of cases (e.g., Bollen and Jackman 1985).

In contrast, just as was true for specific pieces of evidence, qualitative scholars do not necessarily treat all cases as equal; some cases are more "important" than others. For example, in the qualitative tradition, researchers explicitly pursue "most likely," "least likely," and "critical" case study research designs (Przeworski and Tuene 1970;

Collier 1993; George and Bennett 2005). These kinds of research designs assume that the research community has prior theoretical knowledge that makes certain cases especially interesting and important when testing theory.

In addition, because qualitative researchers are interested in individual cases, they are aware of and concerned with cases that are regarded as substantively important. Here "substantively important" means of special normative interest because of a past or current major role in domestic or international politics. For example, qualitative scholars might have serious doubts about a theory of American elections that failed miserably for California and New York even if it worked well for some smaller states. To consider an international relations example: scholars in security studies believe that the ability of realism to explain the end of Cold War is absolutely crucial. For some social constructivists, in fact, a failure of realism to explain this single case represents a major strike against the whole paradigm. Realists seem to agree and work hard to explain the end of Cold War (there is a massive literature on this debate; see, for example, the exchange between English 2002, and Brooks and Wohlforth 2000, 2002). Our view is that qualitative researchers almost instinctively understand the requirement of getting the "big" cases right and worry deeply when it is not met.

The general point is nicely illustrated with an example from Goldstone (2003). Goldstone explores the consequences for Marxist theory of a failure to adequately explain the French Revolution: "It might still be that the Marxist view held in other cases, but finding that it did not hold in one of the historically most important revolutions (that is, a revolution in one of the largest, most influential, and most imitated states of the its day and frequent exemplar for Marxist theories) would certainly shake one's faith in the value of the theory" (2003, 45–46). Within quantitative framework, by contrast, the French Revolution does not count extra for falsifying theory. If many other cases conform, the nonconformity of the French Revolution is not a special problem (or at least no more of a problem than, say, the Bolivian Revolution would be).

The qualitative concern with important cases is puzzling for a quantitative scholar. There is no real reason why substantively or historically important cases are the best ones when evaluating a theory. It could well be that an obscure case has the key characteristics needed to test a theory. In addition, it is unclear why important cases should "count" for more in evaluating theories. If theoretical and empirical scope statements are important – which we believe they are in both qualitative and quantitative research – then it would be better to explain more cases than to evaluate the theory primarily against what might be very idiosyncratic important cases.

9 Lack of fit

In qualitative research, the investigator is quite familiar with each case under investigation. As a consequence, a particular case that does not conform to the investigator's causal model is not simply ignored. Instead, the researcher seeks to identify the special factors that lead this case to follow a distinctive causal pattern. These special factors may not be considered part of the central theoretical model, but they are explicitly identified and discussed. The qualitative researcher therefore seeks to understand exactly why the particular case did not conform to the theoretical expectation (Ragin 2003, 135–38).

By contrast, in quantitative research, the failure of a theoretical model to explain particular cases is not a problem as a long as the model provides good estimates of parameters for the population as a whole. Many idiosyncratic factors may matter for particular cases, but these factors are not important for more general theory, and therefore they are not of great concern.[18] The exclusion of idiosyncratic factors does not bias the parameter estimates of the model given that these factors are often not systematically correlated with error terms specified in the model. Moreover, the lack of fit of a theoretical model may be due not simply to omitted variables but also to randomness and nonsystematic measurement error – problems which again do not bias results.

These different approaches to dealing with a lack of fit provide ample ground for misunderstandings. Qualitative researchers believe that prediction error "should be explained, rather than simply acknowledged" (Ragin 2003, 138). Given this belief, they may be troubled by statistical models that explain only a small portion of the variation of interest, leaving the rest to the error term. They may ask, "What are the various factors that comprise the error term?" If the overall fit of the statistical model is not very good, they may be unconvinced by the argument that the error term contains only minor variables (or measurement error or inherent randomness). For their part, statistical researchers may be perplexed when qualitative researchers spend a great deal of energy attempting to identify factors at work in nonconforming cases. They may wonder, "Why use up valuable time on research that does not lead to generalizable findings?" Indeed, they may view the effort of fully explaining the outcome of interest as a deterministic trap or a utopian goal.

Yet, we are convinced that when one appreciates the different research goals pursued by qualitative and quantitative analysts, it is hard to condemn or even criticize either viewpoint. If you really want to estimate average causal effects, you should not be in the business of trying to hunt down each little causal factor that

[18] The view of statistical researchers on this issue is nicely captured by King, Keohane, and Verba's one effort to discuss causal explanation for an individual case. The authors describe a research project in which the goal is to evaluate the effect of incumbency on elections. King, Keohane, and Verba realize that other variables particular might come into play, but these are relegated to the error term and are of no particular interest:

> [W]e have argued that social science always needs to partition the world into systematic and nonsystematic components ... To see the importance of this partitioning, think about what would happen if we could rerun the 1998 election campaign in the Fourth District of New York, with a Democratic incumbent and a Republican challenger. A slightly different total would result, due to nonsystematic features of the election campaign – aspects of politics that do not persist from one campaign to the next, even if the campaigns begin on identical footing. Some of these nonsystematic features might include a verbal gaffe, a surprisingly popular speech or position on an issue ... We can therefore imagine a variable that would express the values of the Democratic vote across hypothetical replications of this same election. (King, Keohane, and Verba 1994, 79).

might affect outcomes in particular cases. But if you really want to explain outcomes in particular cases, it makes very good sense to be in this business.

10 Concepts and measurement

It is common in qualitative analysis for scholars to spend much time and energy developing clear and precise definitions for concepts that are central to their research. They do so because they are concerned with the problem of conceptual validity, and they believe that the failure to address this problem is a major source of measurement error. When analyzing multiple cases, these researchers especially try to avoid conceptual stretching, or the practice of applying a concept to cases for which it is not appropriate (Sartori 1970; Collier and Mahon 1993). Debates about measurement validity in this research tradition are therefore often focused on the logical structure and content of specific concepts (see Gerring 2001; Goertz 2006).

In quantitative research, by contrast, the focus is less on measurement error deriving from the definition and structure of concepts. Instead, this research tradition is more concerned with operationalization and the use of indicators. For quantitative researchers, measurement error typically occurs at the level of indicators, not the level of concepts, and methodological discussions of measurement error therefore concentrate on modeling measurement error and modifying indicators with little concern for concept revision. In fact, some (though certainly not all) quantitative researchers would go so far as to say that a concept is defined by the indicators used to measure it, a position that qualitative researchers would almost never endorse.

We can see these differences clearly in comparative research on democracy. In the qualitative research tradition, debates over the (mis)measurement of democracy often focus on the stretching of this concept to cases that are not really democracies (or are special kinds of democracies). Solutions to the problem are proposed at the conceptual level – e.g., developing appropriate subtypes of democracy that will simultaneously allow researchers to capture diverse forms of democracy and avoid stretching the concept (Collier and Levitsky 1997). By contrast, discussions of about the (mis)measurement of democracy in quantitative research are concerned with the properties of indicators and the statistical measurement model, including error (e.g., Bollen 1980; 1993; Bollen and Paxton 1998). It is standard in this research tradition to believe that many measurement problems result from the use of poor or biased indicators of democracy.

We have seen these differences contribute to skeptical views across the traditions. For example, qualitative researchers sometimes believe that the indicators used in statistical research are simplistic measures that omit key elements (or include inappropriate elements) of the concept being studied (Coppedge 1999; Munck and Verkuilen 2002; Bowman, Lehoucq, and Mahoney 2005). They may feel that statistical indicators do not measure the same thing across diverse contexts and thus that there is significant unrecognized conceptual heterogeneity is present in quantitative research.

Yet we believe this skepticism ultimately emanates from assumptions particular to qualitative research; it travels less well when applied to the assumptions of quantitative research. In particular, qualitative researchers are skeptical about many quantitative indicators because their goal is to develop adequate explanations of each particular case, and they therefore really must try to measure all key variables correctly for each case. In the qualitative tradition, in fact, scholars actively discuss and debate the scoring of particular variables for specific cases. The stakes of such discussions may be high, for theory falsification might occur with a change in the value of one or a small number of variables. In qualitative research, in short, measurement error needs to be addressed and eliminated completely, if possible. Indicators that on average do a reasonable job of measurement will be problematic because they will incorrectly measure many particular cases.

For quantitative researchers, by contrast, measurement error is something that is unavoidable but not devastating so long as it can be adequately modeled. If the goal is to estimate average causal effects for a large population, it makes good sense that measurement error per se would not be viewed as that large of a problem. Systematic measurement bias is of course important and procedures exist to identify it (for example, Bollen and Paxton [1998] test whether quantitative data on democracy sets are systematically biased in favor of pro-Western dictatorships). And when systematic measurement error is discovered, quantitative researchers will normally seek out better indicators for the concept being measured or better ways to model error. But nonsystematic measurement error itself is generally speaking viewed as inevitable and not devastating.

Given these differences, it is appropriate to speak of two separate strands in the methodological literature on measurement error in political science: a qualitative strand that focuses on concepts, conceptual validity, and is centrally concerned with eliminating measurement error; and a quantitative strand that focuses on indicators, measurement validity, and seeks to model measurement error and avoid biased error. Both literatures are hugely influential within their respective cultures, but cross-cultural communication between the two is relatively rare (though see Adcock and Collier 2001; Goertz 2006).

Conclusion

Comparing differences in qualitative and quantitative research in contemporary political science entails traversing sensitive ground. Scholars associated with either tradition tend to react defensively and in exaggerated ways to criticisms or perceived mischaracterizations of their assumptions, goals, and practices. The possibility for misunderstanding is manifold.

This paper is not as an effort to advise either qualitative researchers or quantitative researchers about how they should carry out work within their tradition. Nor is it an effort to criticize research practices – within the assumptions of each tradition, the research practices we have described make good sense. We thus hope that scholars will read this paper with the goal of learning more about how the "other side" thinks about research. We especially hope that scholars will not read the paper with the goal of noting how the assumptions of the other side are deeply flawed from within their own culture. Given the different assumptions and research goals underlying the two traditions, it necessarily follows that what is good advice and good practice in qualitative researcher might be bad advice and bad practice in quantitative research (and vice versa). In this framework, it is not helpful to condemn research practices without taking into consideration basic research goals.

Misunderstandings across the two traditions are not inevitable. Insofar as scholars are conversant in the language of the other tradition, and interested in exploring a peaceful and respectful dialogue, they can productively communicate with one another. We hope that our listing of differences across the two traditions might contribute to this kind of productive communication.

References

Achen, C.H. Two cheers for Charles Ragin. *Studies in Comparative International Development* 40: 27–32.

Achen, C.H., and D. Snidal. 2004. Rational deterrence theory and comparative case studies. *World Politics* 41:143–69.

Adcock R, Collier D. 2001. Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* 95: 529–546.

Allison, P. 1977. Testing for interaction in multiple regression. *American Journal of Sociology* 83:144–53.

Angrist, J., and G. Imbens. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90:431–42.

Angrist, J., G. Imbens, and D. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444–55.

Aron, R. 1986 (1938). *Introduction à la philosophie de l'histoire: essai sur les limites de l'objectivité historique,* 2nd edition. Paris: Gallimard.

Ayer, A. 1946. *Language, truth and logic.* New York: Dover.

Beck, N. This Issue. Is causal-process observation an oxymoron? A comment on Brady and Collier (eds.) Rethinking Social Inquiry. *Political Analysis*.

Bollen, K.A. 1980. Issues in the comparative measurement of political democracy. *American Sociological Review* 45:370–90.

Bollen, K.A. 1993. Liberal democracy: validity and method factors in cross-national measures. *American Journal of Political Science* 37: 1207–30.

Bollen, K.A., and R.W. Jackman. 1985. Regression diagnostics: an expository treatment of outliers and influential cases. *Sociological Methods and Research* 13:510–42.

Bollen, K.A., and P. Paxton. 1998. Detection and determinants of bias in subjective measures. *American Sociological Review* 63:465–78.

Bowman, K., F. Lehoucq, and J. Mahoney. 2005. Measuring political democracy: case expertise, data adequacy, and Central America. *Comparative Political Studies* 38:939–70.

Brady, H.E. 2004a. Doing good and doing better: how far does the quantitative template get us? In H.E. Brady and D. Collier (eds.) *Rethinking social inquiry: diverse tools, shared standards.* Lanham: Rowman and Littlefield.

Brady, H.E. 2004b. Data-set observations versus causal-process observations: the 2000 U.S. presidential election. In H.E. Brady and D. Collier (eds.) *Rethinking social inquiry: diverse tools, shared standards.* Lanham: Rowman and Littlefield.

Brady, H.E., and J. Seawright. 2004. Framing social inquiry: from models of causation to statistically based causal inference. Paper presented at the annual meetings of the American Political Science Association.

Brady, H.E., and D. Collier (eds.). 2004. *Rethinking social inquiry: diverse tools, shared standards.* Lanham: Rowman and Littlefield.

Brambor, T., W. Clark, and M. Golder. 2006. Understanding interaction models: improving empirical analyses. *Political Analysis* 14:63–82.

Braumoeller, B. 2003. Causal complexity and the study of politics. *Political Analysis* 11:209–33.

Braumoeller, B. 2004. Hypothesis testing and multiplicative interaction terms. *International Organization* 58:807–20.

Braumoeller, B. This issue. Explaining variance: exploring the neglected second moment. *Political Analysis.*

Braumoeller, B., and G. Goertz. 2000. The methodology of necessary conditions. *American Journal of Political Science* 44:844–58.

Brooks, S., and W. Wohlforth. 2000. Power, globalization, and the end of the Cold War: reevaluating a landmark case for ideas. *International Security* 25:5–53.

Brooks, S., and W. Wohlforth. 2002. From old thinking to new thinking in qualitative research. *International Security* 26:93–111.

Clarke. This Issue. Title. *Political Analysis*.

Collier, D. 1993. The comparative method. In A. Finifter, ed., *Political science: the state of the discipline II.* Washington, D.C.: American Political Science Association.

Collier, D., and S. Levitsky. 1997. Democracy with adjectives: conceptual innovation in comparative research. *World Politics* 49:430–51.

Collier, D., and J.E. Mahon. 1993. Conceptual stretching revisited: adapting categories in comparative analysis. *American Political Science Review* 87:845–55.

Collier, D., J. Mahoney, and J. Seawright. 2004. Claiming too much: warnings about selection bias. In H.E. Brady and D. Collier (eds.) *Rethinking social inquiry: diverse tools, shared standards.* Lanham: Rowman and Littlefield.

Coppedge, M. Thickening thin concepts and theories: combining large-n and small in comparative politics. *Comparative Politics* 31:465–76.

Coppedge, M. Forthcoming. *Approaching democracy*. Cambridge: Cambridge University Press.

Dawid, P. 2000. Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association* 95:407–50.

Dion, D. 1998. Evidence and inference in the comparative case study. *Comparative Politics* 30:127–45.

Drèze, J., and A. Sen. 1989. China and India. In *Hunger and public action.* Oxford: Clarendon Press.

Elman, C. 2005. Explanatory typologies in qualitative studies of international politics. *International Organization* 59:293–326.

English, R. 2002. Power, ideas, and new evidence on the Cold Wars end: a reply to Brooks and Wohlforth. *International Security* 26:70–92.

Esping-Andersen, G. 1990. *The three worlds of welfare capitalism.* Cambridge: Polity Press.

Fearon, J. 1996. Causes and counterfactuals in social science: exploring an analogy between cellular automata and historical processes. In P. Tetlock and A. Belkin (eds.) *Counterfactual thought experiments in world politics.* Princeton: Princeton University Press.

Freedman, D.A. 1991. Statistical models and shoe leather. In P. Marsden (ed.) *Sociological Methodology.* San Francisco: Jossey-Bass.

Geddes, B. 2003. *Paradigms and sand castles: theory building and research design in comparative politics.* Ann Arbor: University of Michigan Press.

Geddes, B. 1991. How the cases you choose affect the answers you get: selection bias in comparative politics. In J.A. Stimson (ed.) *Political Analysis*, vol. 2. Ann Arbor: University of Michigan Press.

George, A.L., and A. Bennett. 2005. Case studies and theory development in the social sciences. Cambridge, Mass.: MIT Press.

Gerring, J. 2001. *Social science methodology: a criterial framework.* Cambridge: Cambridge University Press.

Harvey, F. 2003. Practicing coercion: revisiting successes and failures using Boolean logic and comparative methods. In G. Goertz and H. Starr (eds.) *Necessary conditions: theory, methodology, and applications.* New York: Rowman & Littlefield.

Goldstone, J.A. 1997. Methodological issues in comparative macrosociology. In *Comparative social research*, vol. 16. Greenwich: JAI Press.

Goldstone, J.A. 2003. Comparative historical analysis and knowledge accumulation in the study of revolutions. In J. Mahoney and D. Rueschemeyer (eds.) *Comparative historical analysis in the social sciences*. Cambridge: Cambridge University Press.

Goldstone, J.A., et al. 2000. State failure task report: phase III findings. University of Maryland.

Goldthorpe, J. 1997. Current issues in comparative macrosociology: a debate on methodological issues. In *Comparative social research*, vol. 16. Greenwich: JAI Press.

Goertz, G. 2006. *Social science concepts: a user's guide.* Princeton: Princeton University Press.

Goertz, G., and J. Mahoney. 2006. Scope: causal and conceptual homogeneity in qualitative research. Working paper.

Goertz, G., and H. Starr (eds.) 2003. *Necessary conditions: theory, methodology, and applications.* Lanham: Rowman and Littlefield.

Goodwin, J., and T. Skocpol. 1989. Explaining revolutions in the contemporary Third World. *Politics and Society* 17:489–509.

Hall, P.A. 2003. Aligning ontology and methodology in comparative research. In J. Mahoney and D. Rueschemeyer (eds.) *Comparative historical analysis in the social sciences.* Cambridge: Cambridge University Press.

Heckman, J.J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5:475–92.

Hicks A. M., J. Misra, and T. Nah Ng. 1995. The programmatic emergence of the social security state. *American Sociological Review* 60: 329–50.

Holland, P. 1986a. Statistics and causal inference. *Journal of the American Statistical Association* 81:945–60.

Holland, P. 1986b. Statistics and causal inference: rejoinder. *Journal of the American Statistical Association* 81:968–70.

Honoré, T., and H. L. A. Hart. 1985. *Causation in the law,* 2nd edition. Oxford: Oxford University Press.

King, G., R. Keohane, and S. Verba. 1994. *Designing social inquiry: scientific inference in qualitative research.* Princeton: Princeton University Press.

King, G., and L. Zeng. 2001. Logistic regression in rare events data. *Political Analysis* 9:137–63.

Lieberman, E.S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99:435–52.

Lieberson, S. 1991. Small Ns 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.

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.

Mackie, J. 1980. *The cement of the universe: a study of causation.* Oxford: Oxford University Press.

Mahoney, J. 2000. Strategies of causal inference in small-N research. *Sociological Methods and Research* 28:387–424.

Mahoney, J. 2004. Comparative-historical methodlogy. *Annual Review of Sociology* 30:81–101.

Mahoney, J., and G. Goertz. 2004. The Possibility Principle: choosing negative cases in comparative research. *American Political Science Review* 98:653–69.

McKeown, T.J. 1999. Case studies and the statistical worldview: review of King, Keohane, and Verba's *Designing Social Inquiry. International Organization* 53:161–90.

Moore, B. 1966. *The social origins of dictatorship and democracy: lord and peasant in the making of the modern world.* Boston: Beacon Press.

Munck, G.L., and J. Verkuilen. 2002. Conceptualizing and measuring democracy: evaluating alternative indices. 35:5–34.

Nagel, E. 1961. *The structure of science: problems in the logic of scientific explanation.* New York: Harcourt, Brace & World.

Przeworski, A., and H. Teune. 1970. *The logic of comparative social inquiry.* New York: John Wiley & Sons.

Ragin, C.C. 1987. *The comparative method: moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Ragin, C.C. 2000. *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C.C. 2004. Turning the tables: how case-oriented research challenges variable-oriented research. In H.E. Brady and D. Collier (eds.) *Rethinking social inquiry: diverse tools, shared standards.* Lanham: Rowman and Littlefield.

Sartori, G. 1970. Concept misformation in comparative politics. *American Political Science Review* 64:1033–53.

Schrodt, P.A. This Issue. Beyond the linear frequentist orthodoxy. *Political Analysis.*

Seawright, J. 2005. Qualitative comparative analysis vis-à-vis regression. *Studies in Comparative International Development* 40:3-26.

Shivley, P. This Issue. Case selection: insights from *Rethinking Social Inquiry. Political Analysis.*

Skocpol, T. 1979. *States and social revolutions: a comparative analysis of France, Russia, and China.* Cambridge: Cambridge University Press.

Sobel, M. 2000. Causal inference in the social sciences. *Journal of the American Statistical Association* 95:647–51.

Stokes, S.C. 2001. *Mandates and democracy: neoliberalism by surprise in Latin America.* Cambridge: Cambridge University Press.

Van Evera, S. 1997. *Guide to methods for students of political science.* Ithaca: Cornell University Press.

Vaughan, D. 1986. *The Challenger launch decision.* Chicago: University of Chicago Press.

Waldner, D. 1999. *State building and late development.* Ithaca: Cornell University Press.

Weber, M. 1949. Objective possibility and adequate causation in historical explanation. *The methodology of the social sciences.* New York.: Free Press.

Western, B. 2001. Bayesian thinking about macrosociology. *American Journal of Sociology* 107:352–78.