3

Doing Good and Doing Better: How Far Does the Quantitative Template Get Us?

Henry E. Brady

What kind of contribution is *Designing Social Inquiry* (hereafter KKV) by Gary King, Robert O. Keohane, and Sidney Verba? Consider the traditional distinction between theology and homiletics.

THEOLOGY VERSUS HOMILETICS

Theological seminaries distinguish between theology, or the systematic study of religious beliefs, and homiletics, the art of preaching the gospel convincingly. Theologians ask hard questions, develop new systems of theology, and often espouse opinions that would shock and horrify the practicing and devout members of the religion's congregations. Homiletics is about homilies; it is about sermons that are practical, down to earth, simple, and above all, reliable interpretations of the faith. Religions understand, as the social sciences may not, that the goal is to save souls and not simply to increase our knowledge or understanding of the world. For this reason, both theology and homiletics have pride of place in seminaries.

The social sciences have a great deal of theology, but very little homiletics. Perhaps this is why we have saved so few souls. And it may also be why we do such a bad job of training students. A little homiletics might go a long way toward improving our discipline.

KKV is a homily, not theology. There is art in a good homily. Like all good homiletic literature, KKV puts aside doubt and complexity. After all, who would want to burden the average graduate student with the tedious complexity of St. Thomas Aquinas in *Summa Theologica* or Paul Tillich in *Systematic Theology*? And who would recommend the self-doubt of St. Augustine's *Confessions* or Kierkegaard's *Fear and Trembling* or *The Sickness unto Death*? Better to give them Norman Vincent Peale's *The Power of Positive Thinking*.

KKV, however, is not just about positive thinking. It is closer to Moses Maimonides' *Guide for the Perplexed* or Luther's *A Catechism for the People, Pastor and Preacher*. It has a powerful message about the need for reform, self-sacrifice, and discipline on the part of all political scientists—especially qualitative researchers.[1] It puts forth a simple, straightforward faith. It tries very hard to treat qualitative researchers as souls worthy of salvation. And it envisions a unified social science in which there are "Two Styles of Research, One Logic of Inference" (3).[2] To practice this one logic of inference, KKV presents a simple, unified series of steps, a faith to live by, based upon insights from conventional quantitative methods and econometrics. In chapter 3, for example, we are told to:

- Construct falsifiable theories.
- Build theories that are internally consistent.
- Select dependent variables carefully.
- Maximize concreteness.
- State theories in as encompassing a way as possible.

---

1. *Designing Social Inquiry*, subtitled *Scientific Inference in Qualitative Research*, begins by discussing the relationship between quantitative and qualitative research, but another dichotomy also runs through the book. Quite often the authors are more concerned with juxtaposing "small-N" versus "large-N" research than with the qualitative-quantitative distinction. These are not the same things. Small-N research is often qualitative, but it need not be, and large-N research can be qualitative. Roughly speaking, the qualitative-quantitative distinction revolves around issues of concept formation and measurement whereas the small-N versus large-N distinction brings up problems of defining the relevant populations, sampling from them, and dealing with statistical variability. I argue later in this chapter that these statistical issues are dealt with much more clearly in KKV than are those regarding concept formation and measurement. We return to these issues in chapter 12 below.

2. This phrase resonates especially well with someone like myself who was brought up as a Catholic where the faithful must deal with the mystery of three manifestations of God (in the Father, Son, and Holy Spirit) in a monotheistic religion. By childhood training, I am quite receptive to a message of monomethodism, even in those circumstances where it requires a leap of faith.

In homiletic literature, exhortations such as these should be simple, and they need not always be completely consistent (witness the last two rules listed above). A good sermon should have clear points; it should avoid doubt; it should provide plenty of examples. The goal should be to convert the heathen qualitative researcher to the true faith.

This book—to its credit—does these things. It is an extraordinarily good piece of homiletic literature and it should be used in the classroom. It is very nicely written. It is generally lucid and well organized. No one can fail to hear its message.

And indeed, we should all hear the message that is preached. I, for one, have great sympathy with this enterprise, having spent far too many hours listening to talks on comparative politics in which dependent variables or independent variables (or both) did not vary, in which selection bias seemed insurmountable, in which explanations seemed more like good stories than hard-won insights gained from ruling out alternative possibilities. In my introductory statistics classes, I, too, have tried to point out to comparativists that they could do so much better if they avoided omitted variable bias, stopped selecting on the dependent variable, and so forth. I have used some of the same diagrams displayed in the text of KKV (e.g., figures 4.1, 5.1, and 5.2) to make didactic points about good research.

Why, then, do I find myself worried about what this book tries to do? Perhaps I am worried because, despite the authors' desire for a unified approach to social science, there may be something wrong with quantitative researchers[3]—who luxuriate in large numbers of observations and even the possibility under some circumstances of doing experiments—trying to impose a code of conduct, a morality, taken from their own experiences. Certainly the authors, three of the most distinguished and intelligent political scientists in our discipline, mean well, think well, and write well. But I worry that, in the end, they are a little like the Reverend Ike who, when asked how he reconciled living in luxury while he preached to the poor, responded that he believed that the best thing you could do for the poor was not to be one of them. The book ends, in fact, with a chapter on "Increasing the Number of Observations."[4] Is this the best thing we can do for qualitative researchers: to recommend that they not be "small-N" researchers?

Qualitative researchers may indeed profit by increasing the number of

---

3. Keohane is not a quantitative researcher, but two of the authors, King and Verba, certainly are, and the book's approach is so rooted in quantitative research that it seems fair to make this assertion.

4. This chapter means more and does more than just suggest that qualitative researchers get more data, although that is one of the recommendations. I make more comments about this interesting chapter later in the review.

observations, and one of the great strengths of KKV is that it tries to indicate how the poor in observations can become richer in their understanding. At the same time, the book's unspoken presumption that qualitative researchers are inevitably handicapped by lack of quantification and small numbers of observations is bothersome. It ignores the possibility that quantitative researchers may sometimes be handicapped by procrustean quantification and a jumble of dissimilar cases.

DESCENDING FROM THE
RHETORICAL HEIGHTS

I have a number of specific concerns about KKV. Here I will focus on two: my belief that KKV is handicapped by a view of causality too closely tied to the experimental method, and my desire to see more discussion of measurement problems.

Before addressing these concerns, I wish to establish a fair standard for evaluating KKV. Given that I consider KKV to be a homily, and not a work of theology, it may be worth remarking that the value of the *Baltimore Catechism* in which I was drilled as a child should *not* be measured by its logic and argument. Rather it should be evaluated in terms of how many children it saved from perdition. In the end, I think that is how KKV should be judged. Does it work in a classroom? Does it make us better social scientists? By opening up a dialogue with qualitative researchers, the book does make us better, but in its treatment of causation and measurement, KKV may not help us very much.

Explanation and Causality

After a useful discussion of descriptive inference or "establishing facts" in chapter 2, KKV goes on in chapter 3 to discuss "Causality and Causal Inference." As far as I can tell, they equate explanation with causal thinking.[5] Yet

_____

5. It is not exactly clear how "explanation" fits into KKV's categories of descriptive and causal inference, but one reasonable interpretation is that the authors consider explanation to be identical with causal inference. In the first three paragraphs of chapter 2, they repeatedly refer to the "dual goals of describing and explaining" (34). They also note that "description and explanation both depend upon rules of scientific inference. In this chapter we focus on description and descriptive inference" (34). This suggests that chapter 3, on "Causal Inference," is about explanation. Yet, things cannot be quite so simple, because they go on to say that "as should be clear, we disagree with those who denigrate 'mere' description. Even if explanation—connecting causes and effects—is the ultimate goal, description has a central role in all explanation, and it is fundamentally important in and of itself." The first part of the sentence seems to define explanation as "connecting causes and effects," but the second part seems to say that description is also a form of explanation. In

philosophers of science are not so sure that the only kind of explanation involves causality. Take, for example, "classification" explanations such as the observation that iron has certain properties because it appears in a certain column of the periodic table. This does not appear to be a causal explanation.[6] It could be argued that Bohr's atomic theory and its extensions in modern quantum mechanics provide a causal explanation, but this only amounts to saying that there may be causal explanations as well as classification explanations. Moreover, there was a substantial period of time when the classification explanation was all we had. Should we discard these explanations, even when they are all we have, because they do not appear to be causal? We are not so rich with explanations in the social sciences that we can afford to do this without good reason. Qualitative social scientists, in fact, seem especially fond of typologies and classification systems. Do these tools contribute to the explanatory enterprise? I do not personally have an answer to my question, so perhaps I should not fault KKV for failing to include a discussion of this difficult issue. But it is perplexing and thought provoking.

The approach to causality advanced in KKV is based upon an interesting framework developed by the statisticians Donald Rubin (1974, 1978) and Paul Holland (1986). The great strength of this approach, to my mind, is that it emphasizes that a definition of causality requires (a) the careful description of a counterfactual condition (what would have happened if the cause had been absent?) and (b) a comparison of what did happen with what would have happened had the cause been absent. These are two powerful points, and KKV is to be commended for bringing them to the forefront of our discussion. Researchers of all stripes should spend more time describing the counterfactual world that underlies their "becauses." What does it mean, for example, to say that "turnout is lower in that district because it has a high proportion of minorities"? What is the counterfactual

---

the sentence after this one, KKV retains the duality of description and explanation and seems to equate explanation with causal inference, but the book argues for the primacy of inference over either one: "It is not description versus explanation that distinguishes scientific research from other research; it is whether systematic inference is conducted according to valid procedures. Inference, whether descriptive or causal, qualitative or quantitative, is the ultimate goal of all good social science" (34).

6. For more discussion of this example and whether there are noncausal explanations, see Achinstein (1983: chap. 7). Brody and Grandy (1989) provide an excellent set of readings on these topics. Gary King has suggested (personal communication) that classification is a form of descriptive inference, but this seems to stretch KKV's concept of descriptive inference beyond distinguishing "the systematic component from the nonsystematic component of the phenomena we study" (56). It also adds to the confusion noted in the preceding footnote.

world in which turnout would be higher? Is it simply one with a lower proportion of minorities? Would these nonminorities be like minorities in every other respect except race? How could this happen? What would it mean to have it happen?[7] These are not easy questions.

I have already argued that there might be explanation without causality. I think there might also be causal effects without (much) explanation. Suppose we find, to use KKV's example, that incumbent legislators do better in elections than nonincumbent legislators. Suppose, in fact, we are as certain as we can be about this because we have done an experiment (random term limits, for example) with a large N to test it out. This finding immediately leads to other questions about what aspects of incumbency create this advantage (see, for example, Cain, Ferejohn, and Fiorina 1987). These questions amount to a desire to further specify the causal mechanism. KKV is not averse to specifying causal mechanisms, and the authors say that "any coherent account of causality needs to specify how the effects are exerted," but they believe that "our definition of causality is logically prior to the identification of causal mechanisms" (85–86). This claim of logical priority may or may not be true (I am not sure it is very important), but what is true is that a discussion of causality is inevitably tied up with a discussion of explanation, theories, and causal mechanisms, and KKV does not pay enough attention to this relationship. There is no discussion of Hempel's (1965) covering laws, of Wesley Salmon's (1984) model of statistical explanation, of Scriven's (1975) "Causation as Explanation," and many other important works on this topic. This is surprising because the philosophical literature, at least, cannot seem to separate the discussion of these issues.[8]

The statistics literature, in fact, is exceptional in defining causality without discussing explanation. Perhaps this is because statisticians want a method of inference that relies only upon the research design and the data, and not at all upon the substance of the research. Yet the net result of the Rubin-Holland papers is a definition that seems surprisingly distant from

---

7. I have deliberately chosen an example in which the putative cause is a characteristic that might be thought unchangeable. Holland, for example, argues that it is impermissible to call race or gender a cause because "for causal inference, it is critical that each unit be potentially exposable to any one of the causes. As an example, the schooling a student receives can be a cause, in our sense, of the student's performance on a test, whereas the student's race or gender cannot" (Holland 1986: 946). This point is not much in evidence in KKV, and I think the authors were wise to minimize its importance because it certainly seems possible to imagine a world in which gender or race changes, but nothing else.

8. Brody and Grandy (1989), for example, link them in part 2 of their reader entitled "Explanation and Causality." Scriven (1975) joins the two concepts in his famous article on causation as explanation, and every philosophical writer of whom I am aware deals with explanation and causation together.

the problems of theory building and explanation as it exists in the sciences. Most importantly, this approach provides no guidance on what constitutes a "good" explanation beyond what constitutes a good causal inference. Yet an analysis of the impact of incumbency may be an excellent causal inference while being a bad explanation.[9]

After defining causality, KKV goes on to describe a method for causal inference. In this, as in its definition, KKV is guided by the work of Rubin and Holland. The major strength and weakness of this approach is its reliance upon the metaphor of the controlled experiment for solving the problem of causal inference. Holland tells us that:

> because experimentation is such a powerful scientific and statistical tool and one that often introduces clarity into discussions of specific cases of causation, I unabashedly draw on the language and framework of experiments for the model for causal inference. It is not that I believe an experiment is the only proper setting for discussing causality, but I do feel that an experiment is the simplest such setting. (1986: 946)

Fair enough. But it is worrisome that Holland finds it "beyond the scope of this article to apply the model for causal inference to nonrandomized studies" (949). Holland cites other literature (Rubin 1978) that essentially concludes that nonrandomized studies are exceptionally difficult to analyze. It is telling that Rubin's extension of the basic framework requires modeling "(1) the prior distribution of the potentially observable data, (2) the mechanism that selects experimental units for exposure to treatments and assigns treatments, and (3) the mechanism that chooses values to record for data analysis" (Rubin 1978: 35). This is a lot of modeling, and it only seems possible if we have strong theories to draw upon.

KKV provides a simplified version[10] of the Rubin and Holland framework, and in the process ignores some of its subtleties. The crucial part of

---

9. If the incumbency example does not persuade, consider a doctor called upon to explain the incidence of psychedelic experiences in a remote culture. In an experiment, the doctor shows that a treated group eating a plant diet consisting of peyote, hemp, beans, carrots, and other plants has a statistically significant increase in their incidence of psychedelic experiences. Thus, eating plants causes psychedelic experiences. This is clearly an incomplete explanation. I wish KKV had discussed by what method I might improve it. I think a discussion of a "good explanation" that went beyond methods for finding causal impacts would have gone a long way toward solving this problem.

10. The authors do add one complexity by making a useful distinction between "realized causal effect" and "random causal effect," but they suppress so much notation and philosophical discussion in their presentation that many of the nuances in Holland's (1986) presentation are lost and none of the extensions in Rubin (1978) are discussed.

KKV's argument is its discussion of "Conditional Independence" (KKV 94–96). In the Rubin-Holland setup there are as many dependent random variables as there are variations in the treatment condition or the explanatory variable(s). In the simplest case with two levels of the treatment, this implies two random variables. One describes the values on the dependent variable Y for the situation where all cases in the population[11] get one level of the treatment (call this Y^I to match KKV's terminology) and the other is for the values on the dependent variable for the situation where all cases in the population get the other level of the treatment (call this Y^N and assume for simplicity that it is no treatment at all). In the real world and for any feasible design, at least one of these values must be censored for each case. That is, we cannot give a case some treatment and no treatment at the same time. But Y^I and Y^N are not the censored variables; they include the unobserved (and unobservable) values as well as the observed ones. A reasonable definition of the causal effect of the treatment is the average of Y^I minus the average of Y^N, but this quantity cannot be calculated because of the unobserved values in these two random variables.

In the Rubin-Holland framework, a necessary assumption for estimating a causal effect is independence between the assignment of treatments and the random variables Y^I and Y^N. This ensures, for example, that people who are high on Y^I are not more likely to get a high level of treatment than those who are low on Y^I. Consequently, we can be sure, for a large enough sample, that the size of the causal effect is the difference between (a) the average of the dependent variable for those who did get the treatment (this quantity can be calculated) and (b) the average of the dependent variable for those who did not get the treatment (another calculable quantity). One way to achieve this kind of independence is to have correctly carried out randomized experiments.

KKV's discussion of this is a bit opaque, and the authors seem to conflate the independence assumption with conditional independence.[12] Condi-

---

11. In this exposition I ignore sampling problems by assuming observations are available on all members of the population. If the entire population cannot be observed, then some assumption has to be made about random sampling.

12. This accounts for the confusing set of sentences at the beginning of section 3.3.2 where KKV first says that "conditional independence is the assumption that values are assigned to the explanatory variables independently of the values taken by the dependent variables" and then goes on to say, "that is, after taking into account the explanatory variables (or controlling for them), the process of assigning values to the explanatory variable is independent of both (or, in general two or more) dependent variables, Y_i^N and Y_i^I" (94). The first quoted sentence must refer to the independence assumption (because conditional independence does not assume that the values assigned to the explanatory or control variables are independent of the values of the dependent variables) whereas the second quoted sentence appears to be about conditional independence.

tional independence is the assumption that the values of Y^I and Y^N conditional on "pre-exposure" or "control" variables are independent of the assignment of treatments. This is implied by independence but it is a much less stringent assumption. It is the assumption that is usually required for the analysis of quasi-experiments (Achen 1986). The conflation of these two different assumptions creates difficulties in the exposition because, whereas we have a method of random assignment to treatment for attaining independence, we have no comparable method for ensuring that the conditional independence assumption holds outside of a randomized design. The best we have is the checklist of "threats to internal validity" developed by Donald Campbell with Julian Stanley and Thomas Cook (Campbell and Stanley 1963; Cook and Campbell 1979).[13] The rest of KKV can be considered another approach to developing a checklist of threats to validity.

Unfortunately, KKV does not allow itself enough pages in this short section to make this very important transition from a discussion of causal inference for experiments to causal inference with "quasi-experiments." I wish the authors had taken more time to explain the independence assumption in detail and to show how randomized experiments might provide us with an operational procedure that would make this assumption plausible. In doing this, they would no doubt have come to the conclusion presented by Cook and Campbell (and updated and expanded recently by Heckman 1992) that there are many reasons to worry about the efficacy of randomization when humans are involved. There are numerous ways in which human beings can make the treatment endogenous by changing their behaviors. There are additional problems when dropouts (and hence censoring of observations) vary by treatment. And there are the difficulties of truly randomizing units when they are people or groups. Once these problems are recognized for randomized designs, it becomes easier to understand how difficult it is to ensure conditional independence for nonrandomized designs.

This transition section might also benefit from a more careful discussion of how theories provide the fundamental basis for making a claim of conditional independence. This is an extraordinarily important step, and knowing how to do it can help researchers avoid the inferential nihilism that has crept into some statisticians' discussion of causal thinking in the social sciences (e.g., Freedman 1991). According to this line of thinking, randomized experiments are practically the only reliable way to be confident that

---

13. I was surprised to find that none of Campbell's publications were referenced in KKV. Besides the books referenced in the text of the present chapter, Campbell's selected papers on *Methodology and Epistemology for Social Science* (1988) make excellent reading.

the conditions for reasonable inferences are met. Conditional independence is considered a chimera—seldom justifiable and usually accepted by the researcher as a matter of pure faith and nothing more. Indeed, if I accepted a notion of inference as bare of theories and the logic of explanation as that proposed by Rubin and Holland, I might also be skeptical of conditional independence. But I believe it is possible to use our prior knowledge, our theories, to carry out the three modeling steps laid out by Rubin (and cited above). Hence, I am more sanguine about the possibilities for cautiously asserting conditional independence.

It might be argued that I brood unnecessarily over technical points. But the section on "Conditional Independence" is the linchpin of KKV. The book wants to show us that concepts from conventional quantitative methods and econometrics will improve our ability to do qualitative research. It argues that the essence of good social research is establishing causal effects. This, in turn, requires making an assumption about conditional independence. This assumption, the authors believe, can be made plausible by avoiding clear-cut violations of it described in the statistics literature. Yet at the crucial transitional moment the argument seems muddy to me. Exactly how can we rule out the violations identified by quantitative researchers? Do quantitative researchers do a good job in this regard? How sure can I be that conditional independence holds after I have followed the instructions in KKV?

The authors of KKV go on to make many useful observations about causal assessment (although, to be honest, I think that Donald Campbell and his collaborators have more useful lists of threats to validity and more trenchant comments about the problems of doing quasi-experimental research). However, in KKV the crucial argument about assessing causation seems to be missing.

Measurement

KKV devotes eighteen pages to measurement (151–68). About five pages cover the "nominal, ordinal, interval" distinction found in the classic papers by S. S. Stevens (1946, 1951), and the remaining thirteen are about systematic and nonsystematic measurement error. The major results on measurement error are the classic ones dating from at least Tintner (1952) on how error in the dependent variable does not bias regression results whereas error in the independent variable produces bias in regression coefficients—in fact, biases them unambiguously downward in the bivariate case. These are well-known results, often repeated in one form or another in classic primers on research design such as Kerlinger (1979), but I do not think they get at the heart of what can be learned from the extensive literature on measurement.

KKV probably gives such short shrift to measurement because the authors believe that causal inference, roughly what Cook and Campbell call "internal validity," is the central problem of doing good social science. I trace this belief to their decision to equate explanation with causal thinking, and to define causal thinking in terms of a narrow analogy to the experimental model. Through this progression, the problems of theory construction, concepts, and measurement recede into the distance. Yet it seems to me that concept formation, measurement, and measurement validity are important in almost all research and possibly of paramount importance in qualitative research. Certainly notions such as "civil society," "deterrence," "democracy," "nationalism," "material capacity," "corporatism," "group-think," and "credibility" pose extraordinary conceptual problems just as "heat," "motion," and "matter" did for the ancients. It may be comforting for the qualitative researcher to know that the true effects of these error-laden variables are even larger in magnitude than what we would estimate using a standard regression equation, but most qualitative researchers are struggling with much more basic problems such as figuring out what it means to measure their fundamental concepts. These problems are certainly not solved by telling us to decide whether the concept is nominal, ordinal, or interval and by admonishing us to "use the measure that is most appropriate to our theoretical purposes" (KKV 153).

I will not pretend to have the answers to the problems of measurement validity in qualitative research, but I think that the debates on these problems would have been advanced by citing some of the more recent literature in this area. Among the notions that come to mind, let me mention three topics that might have been included. Something might have been said about the conceptualizations of measurement developed by Krantz et al. in their magisterial three-volume work on *Foundations of Measurement* (1971–1990), the related notions put forth by Georg Rasch in his quirky but very influential work on *Probabilistic Models for Some Intelligence and Attainment Tests* (1980 [1960]), and the fascinating *Notes on Social Measurement* (1984a) penned by Otis Dudley Duncan, who followed up this broadside on the limitations of social measurement with a brief for using Rasch models in the social sciences (1984b). These works show that *qualitative* comparisons are the basic building blocks of any approach to measurement, thus bridging the "quantitative-qualitative" divide by showing that the two approaches are intimately related to one another. This discussion would have easily led to a second topic: the dimensionality of concepts, the nature of similarity judgments that often underlie concept formation, and the role of taxonomies and classifications in science. Finally, there might have been a survey of how the LISREL framework (Bollen and Lennox 1991), especially when it is combined with the "multitrait-multimethod

approach" of Campbell and Fiske (1959), sheds light on the practical problems of measurement.

Let me discuss each of these literatures. Duncan's observations on Stevens's scale types are probably the best starting place:

> I conclude that the Stevens theory of scale types, pruned of its terribly misleading confusion of classifications and binary variables with N scales, augmented to take more explicit account of the scales used in measuring numerousness and probability, and specified more clearly so that the examples could be properly understood and assessed, has utility in suggesting the appropriate mathematical and numerical treatment of numbers arising from different kinds of measurement. *Still, a theory of scale types is not a theory of measurement.* And I, for one, am doubtful that any amount of study devoted to either of those topics can teach you how to measure social phenomena, though it can conceivably be helpful in understanding exactly what is achieved by a proposed method of measurement or measuring instrument. (1984a: 154, italics added)

Lest anyone miss Duncan's point, his next chapter is entitled "Measurement: The Real Thing." What is "the real thing"? Krantz et al. (1971–1990) provide the fullest answer to this question, but Duncan provides a more accessible treatment. Measurement, Duncan argues, is not the same as quantification, and it must be guided by theories that emphasize the relationships of one measure to another. Take, for example, that favorite illustration of introductory methods classes, the measurement of temperature. Although the development of thermometry involves a complicated interplay between theory and invention, one of the important milestones was the discovery of the gas law for which temperature is proportional to pressure times volume. Thermometry only began to progress beyond crude ordinal distinctions such as cold, warm, or hot to true interval scales once laws like the gas law made it clear that temperature could be measured by the change in volume of some material under constant pressure.

One of the distinctive features of this way of measuring temperature is that it relies upon a simple multiplicative law, which relates temperature to two quantities that can be "extensively" measured. Extensive measurement refers to the use of the standard millimeter, gram, second, or some other quantity that can be duplicated so that a number of them can be added together ("concatenated") and compared with some object or phenomenon whose length, weight, duration, or other feature is unknown. There is no such standard for temperature, but it can still be measured because it is related to two quantities that can be measured extensively (i.e., volume as length times width times height and pressure as mass times length per time and area squared).

A fundamental difficulty facing empirical social science is the apparent impossibility of developing extensive measurements of many important theoretical quantities. Consider, for example, the notion of utility that is basic to both economics and public choice theory. Utility cannot be measured extensively, but economists avoid this difficulty through an ingenious ploy: They throw utility out of their empirical models by deriving demand curves from the maximization of utility with respect to a budget constraint that consists of the sum of prices times quantities. This produces a demand curve—an equation in prices and quantities—both of which can be measured extensively. This ploy, unfortunately, does not appear to be readily available to political scientists.

The contribution of Rasch (1980 [1960]; see also Andrich 1988) and of Krantz et al. (1971–1990) in their method of "conjoint measurement" has been to show how measurement can be carried out without an extensive measure that can be duplicated and combined: all that is needed is the ability to make *qualitative* distinctions about the amount of each of several variables that are thought to be multiplicatively related to one another. Rasch's method, designed for scoring achievement tests, has the great virtue that it scores both test-takers and the items on the test simultaneously.

All this fancy talk does not provide us with a straightforward way to measure the basic concepts in qualitative social science, yet it does provide us with some clues about how we might go about measuring these concepts. First, it suggests that we have two basic strategies for measurement. We can either try to define a concept extensively (as with length, weight, prices, or quantities) or conjointly (as with achievement tests and subjective probability). Thus we can measure democracy extensively by the fraction of the population enfranchised or by the number of parties, or we can measure it conjointly by using ratings from knowledgeable observers. If we use the second method, as qualitative researchers might be inclined to do, then we might want to think about whether we should scale the raters as well as the countries that are rated. Maybe raters differ in their willingness to call a country a democracy; maybe they even have biases of some sort or another.

Second, this discussion suggests that theories must help to guide the measurement process. In their impressive series of papers on bias in electoral systems, Gelman and King (1994) follow just this strategy with a simple framework for thinking about representation. Steven Fish (1995) also does this (more implicitly than explicitly) in his discussion of the development of civil society in Russia. One of his indicators of civil society is the aggregation of interests by groups, which he describes as the group's "identification of 'cleavage issues' and the formulation of specific goals and agendas [and] . . . the formation of a collective identity, which includes the identification of a membership" (53–54). Although Fish does not provide

a mathematical description of his measure, it could be conceptualized as the degree to which participation or membership in a group is highly correlated with some politically relevant characteristic or cleavage. This amounts to defining this component of civil society as the product of group participation and a politically relevant characteristic—a multiplicative relationship of the sort described by measurement theorists as indicative of true measurement. Fish's approach makes sense partly because it has exactly this form. Hence, measurement theory provides a clear-cut check on when we can say that we have the framework for measuring something.[14]

This approach leads immediately into the next topic I mentioned above. There is a very rich literature on the "topology" of measurement that indicates what is required for single or multidimensional measures; what is required for dimensionality itself; what is required before something is considered the same as something else; and under what conditions objects can be better taxonomized using "trees" or Euclidian space. These methods are now widely used in biology to inform studies of evolution. I suspect that they would be quite useful for the qualitative researcher who wants to trace the evolution of the concept of democracy over time, or the similarities and differences among contemporary democracies.[15] After all, qualitative researchers often spend a great deal of time and effort developing typologies and taxonomies.

Finally, although I often worry about the wholesale use of LISREL in survey research, I think the marriage of factor analysis to simultaneous equation modeling in LISREL has made many researchers more aware of measurement problems. Kenneth Bollen (1993) presents an exemplary use of this technique in his analysis of ratings, developed by three different scholars, of political liberties and democratic rule in countries around the world. By having two concepts in mind, Bollen is able to search for "discriminant" as well as "convergent" validity as Campbell and Fiske (1959) tell us we should do. Bollen allows for the possibility that raters may have

---

14. Gary King (personal communication) suggests that these are points for quantitative researchers and not qualitative researchers because they deal with quantitative measures. Putting aside the fact that a discussion of measurement error or Stevens's scale types assumes the same thing (and the entirety of KKV is based upon the premise that quantitative methods provide lessons for qualitative researchers), it is worth noting that qualitative researchers also engage in comparisons that amount to a form of measurement. Qualitative researchers should know that quantitative research relies upon just the kinds of comparative statements that are at the core of qualitative research. In fact, a discussion of this sort would lead to a conclusion that qualitative and quantitative research are not really different at all.

15. Those interested in these topics should peruse the pages of *Psychometrika* or the *Journal of Classification.* Krantz et al. (1971–1990) also explore many of these issues.

biases, and he finds, for example, that one rater "tends to favor countries in Central America and South America, western industrial nations, and, to a lesser extent, countries in the Oceania region" while providing lower scores for sub-Saharan Africa, Eastern Europe, and Asia. One can imagine extending Bollen's work by adding other methods for rating democracy and by examining (as he does in a preliminary way) how the characteristics of the raters affect their ratings. Bollen's work suggests that qualitative researchers might improve their understanding of concepts by considering various definitions of them, by considering concepts closely related to them, and by considering concepts that are different from them. This strategy, for example, is followed by Hanna Pitkin in her classic work on representation (1967).[16]

An exploration of measurement issues along the lines sketched above would benefit both quantitative and qualitative researchers. Indeed, a discussion of these matters is worthwhile even if it only shows qualitative researchers how quantitative work must also grapple with complex measurement problems. Because its authors want to be constructive and want to instruct, KKV invariably tries to show how quantitative notions can improve qualitative research. This is laudable, but it leads the authors to neglect the multitude of problems that confront quantitative researchers, and it ignores the extent to which quantification is based upon qualitative judgments. Both qualitative and quantitative researchers might benefit from a less didactic approach that revealed problems as well as putative solutions. This might lead to a common effort to solve problems of concept formation and measurement that vex both quantitative and qualitative researchers.

CONCLUSION

KKV is an excellent sermon, without much condescension, on what qualitative researchers can learn from quantitative researchers. As a work on methodology it has some substantial defects, such as equating explanation with causal inference, proposing a narrow definition of causality, and drawing far too little sustenance from a strong literature on measurement and concept formation. But it also has substantial strengths. First and foremost, it opens a conversation between qualitative and quantitative researchers, and that is very good. Second, its presentation of causal thinking in terms of counterfactual reasoning forces researchers to consider more carefully the

---

16. Pitkin, of course, describes her methodology as "linguistic" analysis, and quantitative researchers might improve themselves by becoming more familiar with her methods.

counterfactuals behind their putative causal models. Third, it has an interesting discussion of selection bias that should be useful to many researchers.[17] Fourth, the final chapter on "Increasing the Number of Observations" is one of the most important notions in the book. I wish KKV had given more concrete examples of how to do this, and I wish the authors had warned of the dangers of spatial and temporal autocorrelation that can thwart innovative attempts to increase observations, but the basic concept is a very important one.

Students will definitely profit from reading this book. The discipline has already benefited from the discussions it has kicked off. I look forward to seeing a generation of graduate students uplifted and improved by reciting its useful and informative homilies.

---

17. I wish, however, that they had not used the term "selection bias" (KKV 126) in an example that clearly involves sampling error. The example is presented in a section entitled "The Limits of Random Selection" so the authors may have not meant to use the term "selection bias" except in a colloquial fashion, but it is disconcerting, and certainly confusing, nevertheless.