Evidence

HOWARD S. BECKER

The University of Chicago Press
Chicago and London

CONTENTS

ONE

Models of Inquiry: Some Historical Background

Alain Desrosières (2002) suggested that we think about the development of data, and methods for turning it into evidence, in ways proposed by contemporary work in the sociology of science. He described how the kind of statistical data social scientists use today took shape from the activities of the functionaries of developing European states who needed systematic information so that they could adequately administer the ever-larger territories under their control. And so, unable to get data as accurate as they wanted, they dealt with the resulting uncertainties by developing mathematical methods for estimating the probabilities associated with their conclusions.

Desrosières traces the way modern statistical method and practice developed to do the work whose results these people needed, "the task of objectifying, of making *things that hold*, either because they are predictable or because, if unpredictable, their unpredictability can be mastered to some extent, thanks to the calculation of probability" (2002, 9). The objects so made embody one kind, perhaps the model we all almost instinctively have in mind, of *data*. Their ability to *hold*, to stay constant, is what allows them to work as evidence. When we point to these things-that-hold, we do it confidently, knowing that our scientific peers will agree that those data support the idea we say they support.

Desrosières describes two things researchers have to do to get that kind of assent from their audiences: "On the one hand, they will specify that the measurement *depends on conventions* concerning the definition of the object and the encoding procedures. But, on the other hand, they will add that their measurement *reflects a reality*. . . . By replacing the question of *objectivity* with that of *objectification* . . . reality appears as the product of a series of material recordings: the more general the recordings—in other words, the more firmly established the conventions of equivalence on which they are founded, as a result of broader investments—the greater the reality of the product" (2002, 12). And thus the more convincing they are as evidence. I'm concerned with the work done by the "conventions of equivalence" that let us accept the "reality" of what are after all pretty shaky data (no matter how scientific our methods of gathering them). So, yes, our data rest on an agreement to accept as good enough for our purposes the less than perfectly reliable objects our methods of objectification produce.

Social scientists work under conditions they can't control. Unlike some other scientists, we can't even pretend to be sure that the "all other things being equal" condition, so central to the model of experimental control as a way of isolating causal links, ever holds for the data we gather. We're always contending with events and people who interfere with our plans for collecting data that stands up, "holds," as evidence for our ideas. As a result, skeptics always have a good chance of falsifying the links we make to connect our data, evidence, and ideas. Critics can find reasons to reject the data's value as evidence for the idea presented, claiming that something other than what the presenter claims might have produced the same results, pointing to the possibility of errors of observation, analysis, or reporting. Or they can claim that the evidence, even if acceptable, doesn't logically support the idea, because . . . and then cite a reason not envisioned in the original research design. Or a critic might argue that the idea is logically fallacious or have some other flaw, rendering untenable the entire argument the research aims to construct.

Disciplines vary in how much their members agree on what they will accept as data "good enough" to serve as evidence for the ideas they are supposed to support. We'll see later that natural scientists have plenty of such troubles themselves but (somewhat) more easily find ways to conquer them. In one extreme and not uncommon case, described by Thomas Kuhn in his classic book on scientific revolutions ([1962] 2012), all (or, more likely, most) members of the natural-science disciplines agree on the basic premises their collective work rests on. They have, in the useful term he gave us, a *paradigm*. They agree on what problems they should be trying to solve and what data will provide convincing evidence to support the particular subideas the paradigm generates. They can tell when they're right and when they're wrong.

Kuhn observed that we seldom see any such happy situation in the social sciences, giving as evidence for that conclusion the data he collected observing the small group of social scientists he joined for a year as a fellow at the Center for Advanced Studies in the Behavioral Sciences, a group of some fifty scholars eminent in their various fields: "Particularly, I was struck by the number and extent of the overt disagreements between social scientists about the nature of legitimate scientific problems and methods. Both history and acquaintance made me doubt that practitioners of the natural sciences possess firmer or more permanent answers to such questions than their colleagues in social science. Yet, somehow, the practice of astronomy, physics, chemistry, or biology normally fails to evoke the controversies over fundamentals that today often seem endemic among, say, psychologists or sociologists" (Kuhn [1962] 2012, xlii). These facts, which surprised Kuhn, the physicist turned historian and sociologist of science, infuse the everyday experience of most social scientists, who know from their own work lives that that's just the way people in their fields do things. But they also know that the disagreements vary considerably in degree, permitting enough consensus among at least some of their members that some work ordinarily does get done.

I grew up in a sociological tradition that minimized such conflicts, although it contained plenty of the methodological differences that became more pronounced in later years. The University of Chicago Sociology Department in the post-World-War-II era (approximately the early 1940s until the middle 1950s), still somewhat influenced by the broad and inclusive vision, created and promoted by Robert E. Park, of what sociology could be, harbored all kinds of serious and deeply felt differences of opinion about these matters, but the differences existed—at least this was my experience, and I wasn't the only one—in an atmosphere of general acceptance of multiple ways of doing research on social life. People argued (after all, it was a university department; what else would they do?) about everything but essentially accepted multiple approaches to basic questions, accepted the data their colleagues provided as evidence for their overlapping ideas. Many people utilized multiple forms of data in their studies. Park's students Clifford Shaw and Henry McKay, for instance, studied juvenile delinquency for years using mass quantitative data, generally taken from police statistics and court records, which permitted the use of statistical techniques of data analysis (correlation coefficients, for example). Simultaneously, they studied the same questions in less formalized ways, collecting and publishing detailed life-history materials provided by individual actors, stories of lives in crime, delinquent careers, successes and failures. Others used similar combinations of material to pursue knowledge about the specific experiences that made up criminal careers, suicides, and other such activities. Some of the great community studies of the period—*Middletown* (Lynd 1929), *Middletown in Transition* (Lynd, 1937), *Deep South* (Davis, Gardner, and Gardner 1941), *Black Metropolis* (Drake and Cayton, 1945)—were models of such methodological breadth.

Strong (and stubborn) proponents of differing methodological approaches had major disputes—the disagreements of Herbert Blumer and Samuel Stouffer about what form sociological science should take were legendary—and some people specialized in one method rather than another, but no organized, even institutionalized, conflict went on between what later came to be called "quantitative" and "qualitative" methods. It's true that the building at 1126 E. Fifty-Ninth Street in Chicago, the home of social science at the University of Chicago, bore this legend (attributed to the famous physicist Lord Kelvin) on its facade: "When you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind." But a story lovingly preserved by some of the people who worked in that building, at least in my time, told of the economist Jacob Viner walking by one day, observing Kelvin's remark and saying, contemplatively: "Yes, and when you can express it in numbers, your knowledge is also of a meager and unsatisfactory kind" (Coates and Munger 1991, 275). My introduction to this ecumenical view of my new profession came from Everett Hughes, who had supervised my dissertation. After I got my PhD, the department hired me to do some teaching, which meant that I now attended faculty meetings. I was surprised to see the evident good feeling and friendship between Hughes and William F. Ogburn, who we graduate students (who were not at all aware of what actually went on among faculty members) thought must surely be mortal enemies, and said as much to Hughes. He looked at me like I was insane (I think he must often have had that feeling when I spouted my twenty-three-year-old's opinions) and wanted to know what I was talking about. I explained that we all thought that their evident differences in methods of research must necessarily have created some enmity between them. He humphed and said, "Don't be silly. Will Ogburn and I are the greatest of friends," and then provided what was for him definitive proof: "Who do you think helped me with all the tables in *French Canada in Transition*?" A lesson I never forgot.

Since all our knowledge is unsatisfactory and just a beginning, we shouldn't equate good science exclusively with the kind that uses numbers (or with its opposite) and should instead refuse to add to our troubles in making social science by engaging in that kind of intramural quarreling. Nor should we equate good science exclusively with work whose warrant rests on long immersion in all the details of social interaction and its results as a way to understanding the organization of social life. We can all use the deficiencies in our own way of working as sources of ideas about how to improve our data gathering and evidence-using to generate more and better ideas, which we can then check out with new ways of gathering data, and so on around the circle.

Because data, evidence, and ideas really do constitute a circle of dependencies, we can move in both directions around that circle. We can try the classical route, using data we create as evidence to check out ideas we have already generated. But we can also use data that unexpectedly differ from what we expected, to create new ideas. Depending on the direction you take, you will probably find yourself using different methods of gathering and analyzing data. Both directions work and produce useful results. Some of us will specialize in work going in one direction, seeking ever more accurate ways of measuring to create data that let us test ideas we (or someone else) have already generated. Others will go in the other direction, looking for data whose unexpectedness will provoke new ideas. Some of us will do both, looking for data that let us generate ideas that further our understanding of the social situations we study, and simultaneously working on ways to test the new understandings we have provisionally arrived at. We get further, collectively, by recognizing the multiple ways we can advance knowledge in our field.

I've conceived this book in that spirit, trying to rethink the contemporary split between these two allegedly different ways of doing scientific business, trying to avoid unnecessary quarrelsomeness. And recognizing what's good in every way of working by connecting the variety of methods involved to basic questions about the connection between data, evidence, and ideas. This has led me to revisit a lot of well-known flaws in quantitative work, not to be argumentatively snotty, but to see how recognizing them can be used to improve the way we all do business. And to apply the same serious critical standards to qualitative work as well, identifying flawed procedures and looking for ways to improve them. And, especially, to call attention to the long-standing (though often overlooked) tradition I've already mentioned that combines both kinds of data gathering in the same studies, work that sees and implements the unity in good social science research.

One consequence of reasoning this way is that we can all cultivate flexibility in what we know and what we do, participating and observing at times, counting and calculating at others. Later on, I'll offer examples of excellent research and thinking that proceeded in just that way.

Models of Knowledge

Desrosières, in his masterful history of statistical reasoning (2002), calls attention to two classical models of scientific knowledge, associated with two eighteenth-century scientists, Carl Linnaeus (also known as Linné) and Georges-Louis Leclerc, Comte de Buffon. Linnaeus proposed the use of a fully made classificatory scheme into which scientists could insert the information their research produced. Scientists completed their work when they filled all the slots in the classification scheme with data. Buffon proposed, on the contrary, to make the construction of the classificatory scheme itself the main job to be done, a job that would never end because, he thought, new and unexpected data would continually overflow the then-existing classificatory boxes, requiring rearrangements of ideas into new, until then unexpected, patterns and arguments. Both thinkers investigated animals and plants, but each used the information his research produced in different ways. To repeat, Linnaeus defined the job as slotting research results into the proper boxes in the scheme he had constructed. Buffon saw it as continuing to create new boxes as new facts came to light.

These two modes of analysis differ in their prescriptive forms (but only to a degree) about what research-produced data can and should be used for. Here's Desrosières's analysis of their differences:

> Of all the features available, Linné chose certain among them, *characteristics*, and created his classification on the basis of those criteria, excluding the other traits. The pertinence of such a selection, which is a priori arbitrary, can only be apparent a posteriori; but for Linné this choice represented a necessity resulting from the fact that the "genera" (families of species) were *real*, and determined the pertinent characteristics: "You must realize that it is not the characteristic that constitutes the genus, but the genus that constitutes the characteristic; that the characteristic flows from the genus, and not the genus from the characteristic." . . . There were thus valid natural criteria to be discovered by procedures that systematically applied the same analytical grid to the entire space under study. Valid criteria were real, natural, and universal. They formed a *system*.
>
> For Buffon, on the other hand, it seemed implausible that the pertinent criteria would always be the same. It was therefore necessary to consider all the available distinctive traits a priori. But these were very numerous, and his *Method* could not be applied from the outset to all the species simultaneously envisaged. It could only be applied to the large, "obvious" families, constituted a priori. From that point on, one took some species and compared it with another. The similar and dissimilar characteristics were then distinguished and only the dissimilar ones retained. A third species was then compared in its turn with the first two, and the process was repeated indefinitely, in such a way that the distinctive characteristics were mentioned once and

> only once. This made it possible to regroup categories, gradually defining the table of kinships. This method emphasized local logics, particular to each zone of the space of living creatures, without supposing a priori that a small number of criteria was pertinent for this entire space. . . .
>
> This method is antithetical to Linné's *criterial* technique, which applied general characteristics presumed to be universally effective. (Desrosières 2002, 240–42)

Desrosières saw this difference in method reflected in the daily working problems of social scientists:

> Any statistician who, not simply content to construct a logical and coherent grid, also tries to use it to encode a pile of questionnaires has felt that, in several cases, he can manage only by means of assimilation, by virtue of propinquity with cases he has previously dealt with, in accordance with a logic not provided for in the nomenclature. These local practices are often engineered by agents toiling away in workshops of coding and keyboarding, in accordance with a division of labor in which the leaders are inspired by the precepts of Linné, whereas the actual executants are, without knowing it, more likely to apply the method of Buffon. (242)

Applying his analysis to contemporary sociology shows how these classical differences in aims and procedures produce two somewhat different ways of working that we needn't think of as conflicting but that surely are different in aim and execution.

Linnaeus's Solution

Most conventional training in social science research methods, and most academic procedures surrounding the approval of student research projects and the resulting dissertations, take a ritualized form more often honored ceremonially than in the work students actually do. These formalities, in essence, reflect the idealized Linnaean procedure Desrosières described.

In the purest, most classical form, a dissertation proposal reviews a collection of literature supposed to report on a coherent body of already gathered knowledge that has reached a point where the problem the student proposes to solve represents the next step on the road to an ever-growing system of established, lawlike propositions. I first heard this view of social science expressed in a perhaps apocryphal story about Beardsley Ruml, an economist best known for having invented the idea of withholding taxes from the wages employers paid their employees. Robert Redfield, an anthropologist who taught at the University of Chicago when Ruml was dean of social science there, told a group of us students that the dean had the habit of approaching unwary faculty members and asking, in a booming voice, "What brick have you added to the wall of social science this week?," sometimes varying the metaphor to ask what link the unlucky professor had added to the chain of science. Redfield said he could never think of an adequate answer and preferred, as more realistic, the metaphor of many small streams flowing toward the ocean, some of them occasionally joining to cut a deeper channel.

Proceeding from this necessarily fictitious problem, students detail what everyone else (the "literature") has said about it and then, most importantly, how their research will gather data whose analysis will resolve some existing disagreement and make it possible to adjudicate between rival explanations. Solving the problem the student has proposed as crucial completes the ritual.

But almost invariably, things don't turn out as the proposal suggested they would. The almost always equivocal findings suggest obvious alternative explanations that look just as plausible as the student's proposed hypotheses, and the research ends not with the bang of a definitive yes or no to the hypothesis originally proposed, but with the classical whimper that "further research is necessary."

So that life can go on, and students can receive their degrees, everyone involved agrees to ignore the original proposal and settle for what was actually found and whatever after-the-fact explanation the hapless student cooked up to explain that result.

One kind of research best suits this situation: the more or less classical quantitative research design, which provides the raw material for a substantial proportion of articles found in the major journals of the student's field: a clearly stated problem with an appropriate bibliographical pedigree; a suitable research method, which usually involves a survey carried out on a specified population, analyzed according to some version of what's called the general linear model, in which the researcher tests the effect of several independent variables (individually, and sometimes jointly) on a dependent variable of interest. Studies of the relation between academic or financial achievement, on the one hand, and social class and race, on the other, exemplify this way of working.

These research designs call for what has become the standard method for much of contemporary sociology: amassing large bodies of data, gathered mostly through questionnaires or, alternatively, by using large bodies of information gathered by organizations for their own purposes—censuses; public records of births, deaths, and the causes of deaths; and statistics produced by schools, police departments, hospitals, and all the other organizations that routinely gather such information to tabulate and count. These organizations collect the data for their own administrative (and often quasi-political) use but often allow social scientists to use them for research purposes.

When problems arise in the execution of survey research once designed—difficulties in collecting an appropriate sample of respondents, for instance—researchers can't easily change their plan, because the logic of analysis depends on the plan's appropriate execution, which almost invariably requires a wave of interviews all done at approximately the same time. Otherwise—if interviews are staggered over a lengthy period—intervening events can influence respondents' answers. David Gold, an experienced survey researcher, told me about a survey he administered to two classes of his students at the University of Iowa, one part of which involved their attitudes about the school's football team. One class filled the questionnaire out on Friday, the other on Monday. Over the weekend Iowa's football team enjoyed a great victory, or maybe they suffered a terrible defeat; in any case, the attitudes of the two quite similar populations varied widely depending on the day they answered the questions. The general solution of such problems is to make the newly discovered difficulty the target of a subsequent study (this is standard in, for example, psychological tests of learning theories with experiments on animals).

An alternative version uses data already gathered by others, often a public agency, for their own purposes, but which the researcher can get access to. Classic examples include, for instance, the US Census data and cause-of-death statistics (which provided the raw material for Emile Durkheim's classic analysis of suicide), often offered as a model for this research format. In both cases, once the data-gathering operation begins, it has to be carried out as planned for as long it takes to get it all done. You can't change the method because, no matter what flaws you've discovered in the records from which you extract your data, the data consists of what the people whose working records they are have already made them, and if they embody errors, so be it; what's done is done.

Research done this way has many advantages. It's relatively easy, at least in principle, to cumulate knowledge and add bricks to the wall of scientific understanding of the thing you're studying. Each study nails down some points, adding to the weight of confirming evidence, and exposes some problems, which can be, and sometimes are (though by no means necessarily), taken up and dealt with in succeeding researches.

More specifically, such research can focus on key variables and measure them in a large number of cases, gathering information on hundreds or thousands of people, instead of forty or fifty. As a result, researchers can use complex statistical techniques of analysis and generalize their findings, using probabilistic reasoning, to larger populations of people or cases.

Discoveries come at the end of the process, when you've assembled all the data and can summarize it in arrays, tables, and specific measures. At that point you may well have some new findings to report and make theories from and about. But you can't exploit those findings until you plan and execute the next study.

Buffon's Solution

In an alternative form of research planning and execution, researchers begin with some general, possibly quite vague, guiding ideas about the things they intend to study. Buffon knew there were animals in the world that must be related to each other in some way, but he didn't know how they were related and didn't know whether the categories he had already developed in his work to that point would be adequate to describe and classify new and novel specimens. Because the world likely contained more complicated cases than he knew about, he made it his job to search out the complications and use them to create a still provisional but more adequate classificatory scheme. In the social science version, you start such an investigation with some simple orienting thoughts: where does what I'm interested in take place, who will be there, and what will probably happen—that might be a typical such list. Working this way, the researcher discovers what hitherto unexpected phenomena need understanding and explanation. It's not unlike an anthropologist who ascends the Alto Xingu, the big river in the interior of Brazil, looking for and, with luck, finding a tribal group that hasn't had any previous contact with Europeans; the researcher doesn't know what language they speak or anything about the way they live. Any anthropologist would guess, of course, that such still "uncontacted" people had some kind of kinship system to define and regulate sexual relations and their consequences, some kind of religion to explain things that seemed to have no more practical everyday explanation, and some kind of food-gathering operation—but what kinds of such things aren't yet known. A major part of the work would consist of describing what had to be explained, and this would have to come before any explaining could begin.

W. Lloyd Warner, a social anthropologist who had studied both an indigenous Australian society, the Murngin (1937), and a modern American community, "Yankee City," a pseudonym for Newburyport, Massachusetts (1941–59), gave his students this advice about how to do fieldwork: "When you learn that some major event is going to take place"—an initiation ceremony or a major civic celebration—"you get there before anyone else, stay through the entire event, and be the last one to leave. Then go and talk to everyone who was there and ask them to tell you what happened."

He suggested that if you did that, having shared the event, you would have something specific to ask people about. If you think of social life as a process, as I do—this happens, that happens at the same time, this other thing happens next—you can understand it all better if you find out what "it" is, rather than trying to fit these events into already defined slots.

So in this version of research, the things you learn at the beginning shape, in part, what you look for, what you find that needs explaining. When Blanche Geer and I began our several-years-long study of college undergraduates at the University of Kansas (Becker, Geer, and Hughes 1968), we had behind us a several-years-long study of student culture in a medical school, and had formulated a lot of ideas about how students collaborated to produce bodies of shared understandings about the situation they were in and how to handle it. Clearly we couldn't just transport our conclusions about student culture in the enclosed, narrowly focused, high-pressure situation of the medical school to the quite different situations of an undergraduate college. We had some general orienting ideas—culture grows up among people who share a problematic situation and have opportunities to communicate about the problems it makes for them, for instance. But the situations, so radically different, would probably produce different results, we reasoned, so we couldn't generate any detailed, testable propositions until we knew a lot more than we did when we began our research.

We first set foot on the KU campus, our research site, a few days before classes began, probably during Orientation Week. We were going to study the kinds of shared understandings and organized activities we had discovered in the medical school (Becker et. al. 1961). We wandered around the tables that a variety of student organizations had set up to introduce themselves to the students enrolling at KU for the first time and introduced ourselves to everyone we met as researchers who would be around the campus for the next few years.

One young man, call him "Jack," asked us a lot of questions about who we were, what we were doing, and so on, and then disappeared. He reappeared a few hours later (we learned later that he had spent the intervening time checking on our credentials with a variety of university officials, making sure we were legitimate) and sat us down for a two-hour initiation into some aspects of campus political life that most people didn't know about. He told us, to summarize a long discussion, that there was a secret society on campus, whose members consisted of the leaders of most, if not all, major campus organizations—including the Interfraternity Council, the Panhellenic Organization (women's sororities), and the student government. He said that this group decided, secretly, who would be the next president of this and the next chairman of that, controlled most organizational and political happenings that students could control, and had considerable influence with the top administrators. We charitably decided that he was a little crazy and didn't focus on investigating his ideas.

The next two years of fieldwork taught us that everything he told us was substantially true. Mind you, this was only one among many ideas we were investigating. We had used his bizarre tale, cautiously and with reservations, to orient parts of our inquiry that dealt with those matters, and we kept finding out that things on campus really happened just as he had said they did. We learned one such corroborating item after another, using each one to further orient our inquiries, to frame the questions we asked, to suggest what meetings we should attend, and so forth. We probably would eventually have learned most of these things without his help, but his disclosures speeded the process.

Simultaneously, we used other things we learned to orient other aspects of our work. Here's a striking instance. One afternoon, I sat with two young women who were in their first year at the university and listened while they chatted about this and that. Eventually one of them asked the other about a young man she'd had a date with the previous evening. "What was he like?" "He was really nice, I had a great time. But I'm never going out with him again." "Why not?" "He has a low grade point average [the standard campus measure of academic achievement, the arithmetic mean of grades in courses you'd taken]." The other accepted this as a sufficient explanation of why an otherwise desirable young man would be rejected. I didn't. It didn't sound reasonable to me, with my adult wisdom, so I asked, "What does that have to do with it?" She looked at me pityingly, as you might look at a child who didn't understand some elementary fact of life, and explained, as her friend nodded understandingly, that this was her first year on campus, she would be there for three more years, and she had no intention of getting seriously involved with someone who wouldn't be at the university after this year (assuming, as she did, that he would fail one or more classes and have to leave school). She didn't want to tie herself up socially and romantically in a way that would interfere with her having the kind of good time she expected to have in college. This unexpected observation, among many similar experiences in other areas of campus life, alerted us to the overwhelming importance of the grade point average in student lives. We could never have imagined, well enough to fashion testable hypotheses, that students' romantic lives would reflect the influence of the school's grading system. (In fact, some KU faculty members and administrators we told this story to had trouble believing it.)

Researchers in this style ordinarily do either extensive, long-term fieldwork, sometimes as participants in the activity they're studying, or do lengthy, sometimes unscripted interviews about a common topic. In either case, they use what they learn one day to frame and direct the next day's work. My own work includes both: years of fieldwork, planned from day to day, with musicians and students; and series of detailed interviews for which I created the questions during each interview, tailoring them to the person and the circumstances they were describing to me, with marijuana users, schoolteachers, and people in the world of theater (three different projects, just to be clear). If you work this way, you can quickly reorient your work, incorporating interesting problems you hadn't anticipated into your understanding of the phenomenon you're studying. You ask questions that your first interview provoked in the interviews that follow, and you spend time looking for other instances of an interesting event or idea that can complicate your understanding of it. The research solves some problems and uncovers others in a continuous process, which only comes to an end when time and money and interest run out.

What you can't do is plan ahead in a way that lets you describe what you're going to do to a skeptical audience, such as a dissertation committee or a source of research funding. Nor can you farm the work out to a team of researchers, unless you make them pretty much equal partners in all the work of the project. You never know what your results will be, though you can be pretty sure you'll have some. But neither can you provide definitive proof of anything you want to say, although you can do more in that direction than many fieldwork-oriented researchers do.

Large-scale quantitative research can do the same thing, but the time scale is different. As researchers in this style run into difficulties and sources of error, they can note and report them to their colleagues (as Wallin and Waldo did) and build them into the doing of further surveys and other data-gathering operations. In the end, both kinds of scientists improve their routine procedures and improve the accuracy of their data.

Lieberson's Recommendations

Stanley Lieberson (1992), a distinguished methodologist, has offered a comprehensive description of these two models as they appear in sociological work. Meticulous and scrupulous in his presentation of the models, he eventually concludes firmly that sociologists should use what he describes as a "probabilistic" model. Here's his reasoning:

> On the one hand, at present we assume that evidence that contradicts a theory shows that the theory is "wrong" or at least needs some modification. On the other hand, in the social sciences it is unrealistic to assume that all relevant data will be consistent with a theory even if the theory is correct. Yet evidence in support of a theory is rarely so strong as to eliminate alternative interpretations. Thus, under current procedures, we are damned if we do and damned if we don't. If we are dealing with theories, then we are dealing with evidence. If we take the evidence too seriously, we may reject perfectly decent theories; if we ignore the evidence, we have no theory, merely speculation. How shall we resolve these problems?
>
> The first step is to recognize that we are essentially dealing with a probabilistic world and that the deterministic perspective in which most sociological theories are couched and which underlies the notion of a critical test is more than unrealistic, it is inappropriate. If theories are posed in probabilistic terms, i.e., specifying that a given set of conditions will alter the likelihood of a given outcome, not only will the reality of social life be correctly described, but we will also be freed from assuming that negative evidence automatically means that a theory is wrong. (A deterministic theory posits that a given set of conditions will lead to a specified outcome, pure and simple.) Why is it reasonable to assume a probabilistic rather than a deterministic causal environment? I will ignore the massive, almost infinite array of data errors incurred when we measure social events that may prevent a given result from being observed even if it always occurs. Beyond that, in a complex multivariate world, it is unrealistic to act as if social life is driven by deterministic forces, even if we think it is. Since there is such a wide array of conditions affecting an outcome, it is naive to think that a correct theory will predict or even explain the outcome in any given circumstance. Only the most simplistic and mechanical conception would assume that a theory has to be the dominant influence in all historical settings and contexts, regardless of the heterogeneity of the units. Moreover, a theory that accounted for all events would border on being a history of the world. (7)

This probabilistic approach clearly has a lot to recommend it. Lieberson criticizes modes of research that aim to deal with all the events and objects present in the situations we study, objects and events that he understands perfectly might well influence the outcomes we want to explain:

> From a probabilistic viewpoint, theories incorporating a complex chain of events are unattractive and empirical evidence is likely to be misleading. . . . A theory involving a set of sequences will be useful only if the probabilities within the sequence are all virtually 1.0 and even then the probability value will decline rapidly with the number of sequential events. Suppose the probability of Y, given X, is .7; and the probability of Z, given Y, is .6. The probability of Z, given X, is .7 x .6 = .42. Thus, both theories can be correct, but more often than not a theory that pools multiple steps will make weaker predictions than if each step were looked at as a separate theoretical issue. The errors are even nastier when a second step has a low probability of occurring. For example, to contract paresis, one must first have syphilis, which in turn must progress through several states without treatment. Even then, far less than half of those with untreated latent syphilis get paresis. To be sure, for those with the initial condition (X or syphilis), the probability of the last outcome will be higher than for those not experiencing X or not having a syphilis infection. However, our analysis and understanding are much greater when we examine each part of the chain. Also, there may be parts of complex chains for which we have no theoretical understanding whatever. In the chain of events leading to World War I, what theory accounts for the assassination of Archduke Ferdinand in 1914 and what theory deals with the likelihood of war had there not been an assassination? (8; Lieberson's citation omitted)

But. . . . Yes, there is a big "but" that, for me, changes this sensible assessment. Sociological research need not produce conclusions that can predict the outcomes of a specific set of prior conditions. No need to be able to predict which people will end up with paresis. An alternative goal tells the story of the path that leads to paresis, treating each step as a process to be investigated, a "black box"—more technically, an input-output machine—that contains more complications leading to the end product of paresis. The complications Lieberson finds so troublesome (and he's not the only one) are, for me (I'm not the only one, either), the new things I want to find out about, whose workings I want to incorporate into my understanding of the input-output machine involved in producing paresis. I've gone into the logic of black boxes and their inner working elsewhere (Becker 2014;

see esp. 95–121). Here's a short summary of this position: "We often say, even insist, that social events have multiple causes. But standard methods don't contain mechanisms for searching for causes we don't know about yet. They're good for assessing the degree of the relation between A and B [X and Y in Lieberson's example] but much less good for investigating 'the unexplained variance,' which stays in the black box until we go looking for it" (65). Understanding sociological work as a search for the inner workings of input-output machines, rather than correlations between causes and effects, changes the nature of the enterprise. Working this way, the sociologist looks to add complications to the story rather than to simplify it. This search needn't be antiquantitative, as I've explained at length in discussing such set theoretic methods as quantitative comparative analysis and its analogues in both qualitative and quantitative research (Becker 1998, 183–94). But it doesn't aim to provide verified correlations that can furnish the basis for predictions that can—only provisionally, because the evidence they contain is only probabilistically true—accurately guide the decisions of individuals. Instead, probably more importantly for people who think this way, the goal is to influence the actions of organizations whose managers expect those actions to have verifiable social consequences.

Next. . . .

Both of these models and ways of conceiving and doing research appear continuously in the history of sociology, and a huge literature has grown up assessing their value, their flaws, and the choices that that researchers have to make as they go about their business.

These discussions often have a polemic flavor, insisting, "My way is better than your way." I've done my best to avoid that. Both kinds of research have problems and flaws, and I want to assess them even-handedly, not assign grades, so to speak, but rather see what the problems of research in the sociology business really are and then suggest ways of doing something about them.

Most obviously—I'll give away the easily foreseeable punch line right away—it pays to use both as the circumstances dictate, not taking a quasi-religious attitude toward the difficulties involved in their use, just being practical. We have plenty of examples of excellent research that does that and other examples where the two kinds can contribute in different ways to increasing knowledge. I'm not sure that the "wall of social science" Beardsley Ruml expected will ever be built, but we can do some good work, which would be enough to suit me.

The differences between the two models lie in how each relates to the data-evidence-theory circle. The "quantitative" model most often has trouble with the connection between data and evidence, with showing that the data really measure what the investigator says they should measure to be useful as evidence in the later argument. The "qualitative" model has trouble at the other end, with showing that the collected evidence, though based on acceptable data that are what they claim to be in relation to observed fact, is clearly related to the idea the investigator insists it embodies or demonstrates or is relevant to. Each approach has the advantages it claims, if you allow its premises, but each likewise has characteristic faults it prefers not to deal with unless it has to.

Here's a rough road map of what's to come: a short and selective history of quarrels about methods in sociological research; two examples of what good scientific method consists of, taken from two well-described projects in the natural sciences; then, in part 2, a discussion of the US Census as a prototype of empirical research that raises many of the classic problems, followed by a series of short discussions treating these methodological problems from the standpoint of who actually does the data collection, making the case that the motives, circumstances, and skills of the primary data collectors shape the research results that constitute our data and thus the kind of evidence we can provide for our ideas.