
Writing (and Quantifying) Sociology

Ronald L. Breiger (Sociology)

Undergraduates in my course on quantitative reasoning in sociology soon encounter the homework problem illustrated in Exhibit 1. Having just been introduced to a review of some elementary concepts of statistical "averaging," they are asked to find the average class size of three sections of an introductory course.

Exhibit 1. Textbook Problem: Computing the Mean
At the local university, there are three sections of the basic accounting course, with 8, 12, and 120 students.
A. For the three instructors who teach this course, what average class size do they face?
B. Do the 140 students who are taking the course have a different view? Calculate the average class size that they sit in. (Wonnacott and Wonnacott 1990, 67)

Because sociology in the United States today is famously a numerical enterprise, one that my freshmen, sophomores, and juniors presume emphasizes objectivity of result and accuracy of calculation, and (more directly to the point) since these students have just been exposed to a discussion of the formula for the arithmetic mean, most students decide that part (a) of this textbook problem entails straightforward application of that formula. Thus, the average class size demanded in part (a) is given as the total of students in ratio to the number of sections: $(8 + 12 + 120) / 3 = 140 / 3 = 47$ students per section, rounding to the nearest whole student.

Part (b) of the same textbook problem stumps all but a very few of my
students. Now they are asked about a possible difference in "view" between students and professors. The calculation in part (a) may be seen to have implicitly adopted the perspective of the faculty members who teach the three sections. On the other side of the desk, what do their students experience? Only a tiny fraction (8 of the 140 students considered in this homework problem) are in the smallest class section, the one with but 8 students. At the other extreme, the vast majority (120 of 140) are in the large section of 120 students. On average, the students considered in the homework problem experience a class size of

\[
8 \left( \frac{8}{140} \right) + 12 \left( \frac{12}{140} \right) + 120 \left( \frac{120}{140} \right) = 104
\]

which is more than twice as big a class as the three professors believe they teach.

The arithmetic of parts (a) and (b) of this textbook problem using made-up data is not at all remarkable; no ambiguity in mathematics or logic is uncovered. However, the sociology is profound, and the policy implications consequential, as has been independently discovered by Cornell's dean of university faculty and expounded in a report on class size and curriculum (Cooke 1994).\(^1\) I will emphasize four issues of sociological interpretation that seem implicated in an adequate analysis of this problem in average class sizes. These are considerations that many of my undergraduate students—and not a few of my professional colleagues—find startling.

**Multiplicity.** The same reality requires two different answers in order to compute a statistic as straightforward as the arithmetic mean. The professors on average teach relatively small classes of 47 students who themselves sit (on average) in relatively large classes of 104.

**Reflexivity.** In the textbook example that we have considered, coming to a single correct answer about the mean class size involves the analyst’s placing herself, so to speak, inside the problem. Is one to look from the professors’ “view” or from that of the students?

**Duality.** I propose to use the term “duality” to emphasize the co-constitution of elements existing at different levels of abstraction, as in the principle of geometric duality (Hodge and Pedoe 1968) by which a plane is defined by

---

1. This report noted the discrepancy in class sizes that confronted students and faculty members within the same university, in a manner quite similar qualitatively to that illustrated by the textbook problem of Exhibit 1. Feld and Grofman (1977) provide formal models of the class size paradox.
two intersecting lines, even as a line is determined by the intersection of two planes. In the textbook example considered here, the class size experienced by the students (104 students per professor) is dual to the class size for the professors (3 professors for the 140 students).

Institutional context, or power relations. The answer obtained—whether the class size is small or large on average—depends on which side of the desk is privileged in defining the calculation. As given in the statement of the problem, the instructors "face" a seemingly objective average class size, whereas it is allowed that the students might have a differing "view."

The textbook example helps me to convince my undergraduates that quantitative reasoning in sociology is not a straightforward matter of objectivity or of "letting the numbers speak for themselves," or a process that necessarily entails an omniscient analyst above the fray, or an analyst who could easily be replaced by some computing machinery. Furthermore, these considerations do not (for the most part) lead to any form of nihilistic relativism but, on the contrary, can help the statistical modeler gain a great deal of clarity in describing the world and in comparing alternative accounts. In all these respects, there are deep analogies between writing quantitative analysis and other forms of writing. Issues raised by literary theorists of rhetoric and interpretation such as Steven Mailloux (1989; 1998) can be further illuminated in conversation with quantitative modelers in the social sciences. I will explore some of these analogies in this chapter, focusing on the writing of formulas as an activity that is not at all limited to formulaic writing.

Writing Models

In a thoughtful essay on the relation of sociology and cultural studies, Michael Schudson (1997, 394) urges sociologists to recognize that "one

2. It seems still to be the case that, as Max Weber had it in his 1911 address, "Science as a Vocation" (Weber 1991), “Nowadays in circles of youth there is a widespread notion that science has become a problem in calculation, fabricated in laboratories or statistical filing systems just as ‘in a factory,’ a calculation involving only the cool intellect and not one’s ‘heart and soul.’ First of all one must say that such comments lack all clarity about what goes on in a factory or in a laboratory. In both some idea has to occur to someone’s mind, and it has to be a correct idea if one is to accomplish anything worthwhile. And such intuition cannot be forced. It has nothing to do with any cold calculation. Certainly calculation is also an indispensable prerequisite. No sociologist, for instance, should think himself too good, even in his old age, to make tens of thousands of quite trivial computations in his head and perhaps for months at a time. One cannot with impunity try to transfer this task entirely to mechanical assistants. . . .” This quotation appears at the head of the reading list for my undergraduate course in quantitative reasoning.
of the things we do when we ‘commit a social science,’ as W. H. Auden caustically put it, is to write, and that writing is a socially and culturally constructed act within a cultural field” and not, “as people still like to believe, a transparent record of data or observations.” If as Schudson and others allege there is too little reflexive reflection on the point that “sociology is itself a construction, a writerly construction,” then this absence of self-consciousness would seem to be all the more relevant to sociologists who carry on and report the quantitative modeling and analysis at the core of the field in the U.S. today.

While indeed there is an egregious lack of systematic elaboration of the rhetorical dimensions of quantitative models and analysis, there are also countertendencies worthy of expounding, and openings to a recognition of some points of similarity as well as difference between literary analysis and quantitative analysis of social data. The social scientist who has thought most deeply about issues of rhetoric and quantitative method is an economist, Deirdre McCloskey. In elaborating a rhetoric of statistical significance tests, for example, McCloskey (1998) urges the proposition that good science is good conversation. McCloskey understands that she needs to convince social scientists that elaborating the rhetoric of quantitative analysis is much more than a negative, debunking activity.

The invitation to rhetoric is not, I emphasize, an invitation to “replace careful analysis with rhetoric,” or to abandon mathematics in favor of name-calling or flowery language. The good rhetorician loves care, precision, explicitness, and economy in argument as much as the next person. Since she has thought more carefully and explicitly than most people have about the place of such virtues in a larger system of scholarly values, she may even love them more. A rhetorical approach to economic texts is machine-building, not machine-breaking… an invitation to leave the irrationality of an artificially narrowed range of argument and to move to the rationality of arguing like human beings. (McCloskey 1998, 168)

Within sociology proper there is no elaborated linkage of quantitative modeling and analysis with strategies of rhetoric. There are, however, moves in this direction. Otis Dudley Duncan, a quantitative methodolo-
gist *par excellence* in the American tradition, nonetheless argues for a sociology of social measurement inscribed within reflexivity: "the social roots of social measurement are in the social process itself" (Duncan 1984a, 221). And Duncan has been a tireless critic of

the syndrome that I have come to call *statisticism*: the notion that computing is synonymous with doing research, the naive faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories or the "importance" of the causes of a "dependent variable;" and the delusion that decomposing the covariations of some arbitrary and haphazardly assembled collection of variables can somehow justify not only a "causal model" but also, praise the mark, a "measurement model." (226)

Harrison White, a leading mathematical sociologist and—like Duncan—an elected member of the National Academy of Sciences, has recently argued that, "in avoiding and sidestepping the interpretive—and thus any direct access to construction of social reality—mathematical models have come to an era of decreasing returns to effort." White suggests that "interpretive approaches are central to achieving a next level of adequacy in social data" (White 1997, 58).

One of the problems in furthering conversation between humanists and quantitative sociologists on the subject of rhetoric and method is the tendency of each group of scholars to use second-rank work in castigating the other. It is not uncommon for humanists to imagine quantitative social scientists as modern descendants of Thomas Gradgrind, Dickens’ character in *Hard Times* who carries "the multiplication table always in his pocket, sir, ready to weigh and measure any parcel of human nature, and tell you exactly what it comes to," without acknowledging that we have had quite a few successes in elaborating meaning, values, and social construction in ways that might inform humanistic inquiry. And, for its part,

---

5. Graduate students in the Cornell English department’s radical caucus point out “literature’s disdain for engaging with social science methodologies” (Graduate Student Radical Caucus 1998, 67). A counterexample, an essay by a literary theorist who engages applied social science data analysis with great power, is Ganes (1992). And I would cite Mohr (1998) as one recent review of social science research in support of my contention on the value in overcoming, or at least bracketing, such disdain. Mohr and Duquenne (1997) present and apply quantitative methods for analysis of the sorts of local cultural practices that are at the center of study of at least some theorists of rhetoric (see Mailoux 1989, 145). Griswold (1995) reviews research studies as well as theoretical work in the sociology of literature pertaining to "the relationship between literature and group identities; connecting institutional and reader-response analyses; reintroducing the role of authorial intentionality; and developing a clearer understanding of how literature is and is not like other media."
mainstream sociology seems resolute in not being impressed by "semiotics, or hermeneutics, or any of the other tics of humanistic inquiry in these latter days," often dismissed as "the Paris fashions modish . . . in New Haven and Berkeley and the intellectual suburbs," as one presidential address to a regional sociological society had it (Reed 1989, 2), perhaps in an effort to call attention to the lack of dialogue and the degree to which sociologists are responsible for the absence. Useful conversation will have to step beyond the name-calling.

**Writing Loglinear Models: Changes over the Past Three Decades**

In order to begin a discussion, I would like to offer a reading of some accounts of quantitative modeling in sociology over the past three decades, as a way to understand some of the issues—both practical and general—in writing these formulas. I will refer to models for tables of counted data. Such tables are a stock-in-trade of contemporary sociology; a typical example might be "religious denomination by political party preference." Here we count the number of people in a random sample, taken from a population whose activities we hope to understand better, who are at the intersection of each denomination (perhaps "Baptist," "Presbyterian," and so on) and each party preference (such as "Republican" or "Libertarian"). If people of each religious denomination had the same distribution of party preferences, then we would say that the two variables (religion and party) were "statistically independent," and we could model the expected count in any cell of the table (say, Presbyterian Democrats) by multiplying the number of people in the sample by the overall probability of being Presbyterian and the overall probability of being a Democrat. Notice that this is a multiplicative model, though statisticians and applied data analysts often prefer to replace the term "multiplicative" with "loglinear." Of course in real societies people of different religious denominations do differ in their political party preferences, so the model of simple "statistical independence" that we have been dis-

---

6. Informally, the model is "multiplicative" due to the independence that I have invoked (for illustrative purposes) between religion and party preference. In analogy, the probability that two fair coins both land heads up is \( \frac{1}{2} \times \frac{1}{2} \).

7. The product of terms that are multiplied together is, if transformed in a certain way (by taking its logarithm), equal to the linear sum of the logarithms of the given terms. The "linear" terminology is connotative of a major achievement of statistical science; the formulation of a "general linear model" of which both the usual forms of least-squares regression analysis and "loglinear models" are special cases.
cussing is virtually never adequate to the task of accounting for the observed data. 8

From roughly the 1940s to the 1960s, the inadequacy of the simple independence model was exploited by developing indices of the lack of fit, allowing a sociologist to say for example that religion and party preference were associated however-much more than would be expected in the absence of a relation. While the loglinear model of simple statistical independence did in this way give sociologists a place to stand in assessing statistical relationships, clearly it is not entirely satisfactory to have a single model that the analyst hopes does not fit the data, if one believes that relationships among aspects of social life are important and capable of taking a variety of distinctive forms deserving of distinctive representations in social analysis.

From the mid-1960s through the 1980s, many researchers developed several very large families of generalizations of loglinear models (the model of statistical independence was often a special case of each of these families), the single most prominent contributor to this line of work being a professor of statistics and sociology named Leo Goodman. Historical accounts or glosses on this period of the elaboration of formal models emphasized increasing sophistication, "cumulation" and the production of "verifiable knowledge" (Breiger 1990, 225), and "progress" toward the goal of formulating "relevant models and methods" (Goodman and Clogg 1992 609). This language of cumulation and progress was especially likely to be found in the area of stratification and mobility research. James Coleman (1991, 3-4) contrasted nationally representative studies of individual attainment with an earlier tradition of stratification research based on community probes such as Lynd and Lynd's (1929) study of Middletown. In the earlier studies, analysis of community structure was synthetic, qualitative, concerned with the subjective views of persons located at different points in the structure, and regarded persons within the system as acting purposefully to make use of their resources. In sharp contrast, the more recent national studies of attainment take individual persons as the fundamental units; these studies are analytical and quantitative, provide no subjective views, and regard persons not as purposive actors but—as in much of the research employing loglinear models—as intersections of variables linked together in causal relations. Movement from the first research frame to the second captures for Coleman some as-

8. The "accounting" is done by comparing the observed table of counts to the table that would result by applying the multiplicative or "loglinear" formula reviewed above, where the comparison takes the form of assessment by means of a statistical distribution appropriate to the task of comparing the "observed" and "modeled" counts, the chi-square distribution.
pects of the transformation of social research in general over the past half century. Moreover, Coleman, in contrast to many of his colleagues, was distressed by this line of "progress" (see Breiger 1995 for a review).

Focusing in particular on the development of loglinear modeling in sociology over the past three decades, I find it useful to think of the many different families of models that have been developed as inscribing stylistic differences in writing the structuring of social relationships, with the styles in turn based on different images of the structuring of the cells of a table of data. In introducing each style, I will consider the way in which each path to new models provided a distinctive departure from the simple model of statistical independence. To provide a sense of the images that have been developed by loglinear analysts, consider the following.

**Association models** (Goodman 1984; Duncan 1979). The model of statistical independence introduced above may be seen as the "null association" model in that, if the model fits, Presbyterians and Baptists (for example) do not differ on average in their party preferences; there is zero association between the row categories and the column categories. Other models—models with definite formulas whose degree of fit can be assessed empirically, not just efforts to exploit the lack of fit of the simple independence model—have been developed in which there is postulated to be a "uniform association" across religious groups in party preference (to continue our illustration), or according to which the association is row-specific and/or column-specific (such that, for example, members of a given religious group could be said on average to be particularly strong supporters of a given political party), or perhaps multidimensional. Loglinear models were augmented by log-multiplicative models to handle some of the principal extensions (Goodman 1984). As I have so far presented these matters it all sounds rather straightforward and quite possibly boring; however, by the end of the chapter I will be discussing some major rumblings-up of fiery rhetoric associated with these developments.

**Topological models** (Hauser 1979; Goodman 1984; Erikson and Goldthorpe 1992a). The model of statistical independence consists of a series of interaction terms (one such term for each of the table's cells) each of which is zero. A generalization is to postulate an exhaustive partitioning of the table's cells into sets each of which consists of cells sharing a "constant" interaction parameter (where these estimated constants differ from one set of cells to another). The cells in any "level" may, in principle, be drawn from anywhere in the table.

**Class models** (Breiger 1981, 1990; Goodman, 1984; Marsden 1985; Hout and Hauser 1992). The model of statistical independence is a "one-class" model in that association between rows and columns is homogeneously zero with respect to all rows and all columns. Generalization to a
"class" model results from postulating a rearrangement and partitioning of the rows and the columns of the table into clumps of rows and clumps of columns which produce an exhaustive set of rectangular subtables of the original table, each subtable characterized by the model of statistical independence (exhibiting homogeneously zero association internally) or by a simpler model of homogeneity.

Latent class models (Clogg 1981; Goodman 1987). A given table of counts that has R rows and C columns might not satisfy the model of statistical independence. However, it might be possible to represent the single observed table by a series of tables of imputed counts, each such table also having R rows and C columns, such that each of the constructed tables satisfies the model of simple independence and the imputed counts sum across any cell of the table to the count actually observed in that cell.

Rhetorical models

I have tried to suggest something of the innovation of loglinear modeling by focusing in my brief account on some of the images made available within this modeling framework for distinguishing various possibilities for the structuring of relationships among categories of people such as Presbyterians and Democrats. I would like to move closer now to a rhetorical analysis of these issues, applying a sense of rhetoric influenced by Steven Mailloux’s building on Richard Rorty’s proposal for “a theoretical discourse that focuses not on the confrontation between knower and object—the controlling visual metaphor of traditional epistemology—but on the conversation among knowers” (Mailloux 1989, 144, original emphasis). Following Rorty, Mailloux sees “recent theories of reading” as alternating between two major orientations. A Lockean realism argues that the meaning of a text is independent of the reader, the text consisting of objective facts that are reflected in correct interpretations. On the other side is a Kantian idealism according to which, as Mailloux (144) has it, the minds of readers “constitute the meaningful text,” a text constructed by readers who share interpretive conventions that determine correct interpretations. Mailloux (144–49) argues for “pragmatist readings” according to which

theory soon turns into rhetorical history. More precisely, it becomes a collection of more or less related histories of conversations about texts. These conversations consist of rhetorical exchanges taking place within traditions of arguments and figures, and these traditions of specific rhetorical practices are themselves part of the micropractices that make up a culture at any historical moment. (145).
Quantitative sociologists often flaunt a view of themselves as Lockean realists, and many graduate curricula in methodology are organized around positivist principles. However—and this seems especially true of the leading quantitative researchers and modelers—if one observes these analysts at work, in the midst of carrying out their "micropractices," a much more subtle picture emerges, fraught with ambiguities and prospects for further rhetorical elaboration. To provide some grounding for a pragmatic reading of the literature on loglinear modeling, I will have recourse to the four rubrics that served well to introduce my undergraduates to the subtleties of the "average class size" example: multiplicity, reflexivity, duality, and institutional relations.

**Multiplicity**

As is suggested by my enumeration of different families of loglinear models (association models, topological models, class models, and so forth), there has been a proliferation of loglinear models relative to the ability of a given set of data to help the analyst gain a preference for one model over another. A highly influential early paper of Goodman's, published in 1969 (reprinted in Goodman 1984, 302–41) was entitled "How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables." Such ransacking is evident in that article, for example, in Goodman's discussion (319) of 20 interactions in a table composed of just 9 cells.9

Analyses such as these suggest that our ability to either discover or construct meaning for tables of counted data exceeds the ability of the data to adjudicate between competing interpretive schemes. Panel A of Exhibit 2 (next page) is an excerpt from Otis Dudley Duncan's review of a volume collecting some of Goodman's papers on loglinear models (Duncan, 1982). Panel B of the same exhibit excerpts a gloss on Jacques Derrida's postmodern notion that texts are complex networks of unfinished meanings; for Derrida, therefore, the notion of text is no longer an easy or comfortable one.

In panel A we see Duncan endeavoring to resist the principle of multiple meanings, and we observe as well his difficulties in doing so with re-

---

9. "Ransacking" is a general issue of social science methodology, and it evokes considerable ambivalence. On the other side from many loglinear analysts, one critic who is a professional statistician comments (in an article directed at skepticism toward regression models), "Testing one model on 24 different data sets could open up a serious enquiry: Have we identified an empirical regularity that has some degree of invariance? Testing 24 models on one data set is less serious" (Freedman 1991, 306).
spect to the specifics of loglinear modeling. The first sentence of this excerpt establishes as a clear implication of Goodman's modeling work that models with "radically different conceptual interpretations may be about equally attractive on statistical grounds." The more inventive the analyst becomes in proliferating new families of models within incommensurable families (such that one model ceases logically to "imply" another), the more elusive become the quest for precision and the goal of testing models against data.

We can resolve dilemmas such as this one, Duncan suggests, by focusing our scarce resources for research on "adequate test" of "the really critical hypotheses in a given state of knowledge," pushing for ever larger sample sizes, which would (as the sample size, N, gets larger and larger) in principle allow rejection of any model. But larger N's will not resolve all questions of interpretation—and in some respects they exacerbate them, for if N is big enough we are in danger of rejecting a model even when the data were generated from that model. Hence we confront an "embarrassment" in loglinear modeling, the definition of which is that "the more ingenious we become in devising sophisticated models to reflect the fine points of conceptualization," the more likely we are to be unable to decide which of our models is "true." Greater precision is what leads to greater indeterminacy.

Exhibit 2. On Unfinished Meanings and Interpretive Stress

A. Duncan on Goodman: One of the clear implications of Goodman's examples is that models with radically different conceptual interpretations may be about equally attractive on statistical grounds, so far as one may judge from a particular sample. Thus, on pages 211 and 294 Goodman gives path diagrams for two models fitted to Coleman's two-wave, two-variable panel data on membership in and attitude toward the "leading crowd" for 3,398 high-school boys. In one model there are five (out of the possible six) direct associations linking pairs of observed variables; but in the other model, the observed variables have no direct associations at all, being linked indirectly by way of two latent variables. Both models fit well. Since neither model implies the other, one cannot make an explicit comparison of chi-square values in the manner prescribed for models that are hierarchically related. Nonetheless, both models cannot be "true." It just happens that they give closely similar estimates of expected frequencies in the particular population under study. It would require a much larger sample than 3,398 from that population to achieve rejection of one or both models. Or, one might decide between the
models (or in favor of still another model) on the basis of comparative analyses in a variety of populations. The more ingenious we become in devising sophisticated models to reflect fine points of conceptualization, the more often we must find ourselves with this kind of an embarrassment of riches. The classic “need for additional research” will become increasingly urgent. Perhaps we will be led to a rather different allocation of resources among lines of inquiry such that the really critical hypotheses in a given state of knowledge receive an adequate test. This seldom happens today. A model-rich environment is not necessarily a comfortable one. (Duncan 1982, 961–62)

B. Birch on Derrida: Thus the text, for Derrida, is a complex network of unfinished meanings, a “fabric of traces referring endlessly to something other than itself” [Derrida 1979, 84]. The notion of text is, therefore, no longer an easy, comfortable one, in which openings and endings of meaning are recognizable by readers . . . As a consequence reading becomes a much more dynamic, but uncertain, activity. (Birch 1989, 9)

Duncan would prefer to see this embarrassment as an “embarrassment of riches,” one that urges the researcher on to ever more ingenuity. Duncan recognizes that the success of the loglinear modeling endeavor has led to a surfeit of imposed meanings in the form of models, and he advocates not a formal but a pragmatic approach to living with the resulting indeterminacies, an approach that consists in an (unspecified) interaction between models and data. Duncan concludes the paragraph presented in panel A of Exhibit 2 with the understatement that “a model-rich environment is not necessarily a comfortable one.” There is a useful parallelism between these concerns of the quantitative modeler and those of the literary theorist (glossed in panel B of the same exhibit) who recognizes that the notion of the text is no longer an easy or comfortable one, thus requiring “a much more dynamic, but uncertain, activity.”

The foregoing example demonstrates that some problems of multiplicity in the fit of loglinear models pertain to considerations of sample size. However, other considerations generate similar discomfort. With respect to the topological models introduced earlier, it is possible that quite different specifications of the imputed “levels” of the table lead to the “same” model in that the expected cell frequencies and associated fit statistics produced by each of several models may be identical, to any number of decimal places (Macdonald 1981). Macdonald (562) claims that Robert Hauser, the most prominent researcher associated with the formulation of topological models, sometimes writes as if an assignment of
cells in a table to levels is "self-evident" and that the structure of association in a table may be uniquely captured in the modeling of that table's levels. But such precision of representation is not in general possible, Macdonald claims: "'levels' talk reifies features which are not a necessary component of the model" (562).

Exhibit 3. On Modeling as Narrative and Heuristic

A. Robert Hauser: Up to this point, I have worried along with Macdonald about the possibility of specifying equivalent models, but the fact is that I do not share his concern about the matter. I do not believe that a model consists only of a set of expected values, but it also (and mainly) consists of the structure or story that we use to interpret and explain those expected values. It is all well and good when a model has no equivalents, that is, when it carries unique implications for population data, but that is rare indeed in the social sciences. Most of the time, a model is no more than a vehicle for rendering a complete and internally consistent interpretation of a body of data in light of the ideas we draw from observation, theory, convention, or whatever. Sometimes, the model will illuminate a facet of a phenomenon in a way that we had not anticipated. Sometimes, the fit or findings of one model will lead us to reject other models. In the usual course of social scientific modeling, we are rarely in a position to regard the fit of one model to the data as grounds for rejecting all the other models, but this is the standard against which Macdonald seeks to measure structural models of mobility. We can (and do) learn a great deal from models that fall far short of this goal, and I see no need to propose a special standard for mobility models. (Hauser 1981, 576)

B. Thomas Wilson: [T]he line between mathematical models in some pure sense and the use of mathematics as an aid to data analysis becomes blurred once all mathematical models in the social sciences are recognized as having primarily a heuristic function. For example, one of the major points that has become clear in the rapid development of structural-equation modeling techniques is the importance of correctly specifying the causal structure, and this, it has been repeatedly emphasized in the textbooks, depends on an adequate theoretical understanding of the phenomenon. But exactly the same consideration arises with any mathematical model of social phenomena. To be sure, the models employed in statistical data analysis tend to be closer to the data and more ad hoc than models motivated by
more general concerns, but this is neither an automatic liability nor a guarantee of superiority. (Wilson 1987, 402)

In his reply to Macdonald, Hauser answers (in part) with a highly self-conscious rhetorical position given in panel A of Exhibit 3. Here Hauser claims that modeling the structure of a table is analogous to constructing an interpretive narrative. He acknowledges the innovative point that Macdonald has just demonstrated to his article's readers, who are mostly technical modelers and quantitative analysts: that two quite different topological models might indeed fit a given table exactly as well as each other. But now Hauser asserts that the interpretive force is not confined within the formalities of the models. Models do not consist of their expected values alone or (by extension) the equations that define them precisely; nor is there an easy correspondence between the statistical test of a hypothesis and the judgement we pass on the model generating the hypothesis. In Hauser's formulation in Exhibit 3a model "also (and mainly) consists of the structure or story that we use to interpret and explain those expected values." What are governing are "the ideas we draw from observation, theory, convention, or whatever." The use of "or whatever" seems to suggest a certain distance, however, between Hauser's foregrounded interest in the technicalities of modeling and his more subdued interest in how social meanings might turn into stories that the omniscient "we" have available in some sort of repertoire. Hauser seems to be saying that, although it might be problematic that two different models can't be distinguished from each other, the same isn't true of two different stories; moreover, since models are like stories, Macdonald's critique is not as distressing as it might at first appear.

In this rhetorical moment a leading quantitative methodologist in the mainstream, Robert Hauser, in effect comes close to agreeing with Thomas Wilson, who has endeavored to reconcile interpretive and ethnomet hodological approaches (Wilson 1970) and who writes, in an excerpt reproduced in panel B of Exhibit 3 (from Wilson, 1987), that mathematical models—in general—in the social sciences have primarily a heuristic function. Although they come from different traditions, Hauser and Wilson would seem to agree that interpretation of models is an activity that must take place, at least in part, outside the formal trappings of the models. Both of these highly talented contributors to sociology, in the excerpts of Exhibit 3, depart substantially from the Lockean realism with which Macdonald charges Hauser, with which humanists often charge social scientists, and by which social scientists often define ourselves.
Reflexivity

The notion of reflexivity as developed for example by writers in the traditions of ethnomethodology and conversation analysis draws attention to the fact that "descriptions are not just about something but they are also doing something; that is, they are not merely representing some facet of the world, they are also involved in that world in some practical way" (Potter 1996, 47). The question of whether a given loglinear model is a descriptive account of a table or a theory imposed on it by an analyst is a subtle one implicated in the discussion of the previous section. Even the question of how many rows and columns a table "should" have is an issue that has been brought far beyond the realm of pure description and, so to speak, "inside" the concerns of the formal modelers as an explicitly problematic feature of the analysis that can be modeled in various specific ways. Thus, it is possible to model the combination of row and column categories of a table on the basis of how the resulting combination acts in inducing some sort of an overall structuring among the categories, as in work of mine (Breiger 1981, 1990) and Goodman's (1984; see also Hout and Hauser 1992). These are the "class models" introduced in my earlier fourfold elaboration of loglinear imagery.

Again using Otis Dudley Duncan as my foil, panel B of Exhibit 4 excerpt Duncan's praise of Goodman's work on models for combining categories of tables. I read Duncan's comment as a critique of essentialism and praise for contextualizing the study of statistical relationships within a reflexively defined set of categories that, so to speak, make a given relationship (such as "homogeneity" of mobility flows within and between the categories) work. Emirbayer (1997) and Somers (1998) are elaborating epistemologies of "relational realism" (Somers's term) that aim to supply some theoretical foundations that I see as supporting this kind of modeling. Indeed, once again a parallel with Birch's gloss on Derrida (given in panel A of Exhibit 4) seems warranted.

Exhibit 4. On Classification and Context

A. Birch on Derrida: Following on from Derrida, there is no such thing as the single meaning, the correct meaning, the right meaning. . . . The "rightness" of a decision, of an act of classifying something, is relative not to some inherent correct order for the world ordained somehow in nature, but to a theory, a position, a set of ideas, institutionally created and constructed. (Birch 1989, 25).

B. Duncan on Goodman: Goodman's results make it clear that the classification itself is part of the model specification. Hence it may make little sense to ask, What is "the" relationship between father's
and son’s occupation? Rather, we must ask, What models are consistent with the data in a cross-classification expressed in terms of the following (explicitly defined) categories? (Duncan 1984b, xi)

**Duality**

As I said in my discussion of Exhibit 1, I propose to use the term “duality” to emphasize the co-constitution of elements existing at different levels of abstraction. Duality provides both analytical and rhetorical force to some sociological analyses. The French sociologist Pierre Bourdieu, for example, introduces his graduate students to a “convenient instrument of construction of the object: the square-table of the pertinent properties of a set of agents or institutions” (Bourdieu 1992, 230; original italics). “If, for example, my task is to analyze various combat sports (wrestling, judo, aikido, boxing, etc.), or different institutions of higher learning, or different Parisian newspapers,” Bourdieu (230) tells his graduate students, “I will enter each of these institutions on a line and I will create a new column each time I discover a property necessary to characterize one of them.” This very simple table (in which the number of rows and the number of columns will likely differ, despite its designation as “square”) “has the virtue of forcing you to think relationally both about the social units under consideration and their properties” (230). A well-known example in Bourdieu’s work is a table cross-classifying subjects that would make a beautiful photograph by the occupations of people choosing each subject (Bourdieu 1984, 526). In my terms the “duality” in this example is that categories of people are here defined in terms of their tastes, while the tastes themselves are simultaneously classified on the basis of the occupations of the people who manifest them (see also Breiger 2000). Bourdieu uses tables such as this one to construct “social spaces which, though they reveal themselves only in the form of highly abstract, objective relations, . . . are what makes the whole reality of the social world” (1992, 231).

The quantitative method that underlies the pictorial representations of social fields in Bourdieu’s work is known as correspondence analysis. “If I make extensive use of correspondence analysis,” Bourdieu affirms, “it is because correspondence analysis is a relational technique of data analysis whose philosophy corresponds exactly to what, in my view, the reality of the social world is.” Specifically, “it is a technique which ‘thinks’ in terms of relation, as I try to do precisely with the notion of field” (Bourdieu and Wacquant 1992, 96). Correspondence analysis takes as its material the

---

10. Portions of this section and the next draw directly on Breiger (2000).
"square-table of pertinent properties of a set of agents or institutions" introduced in the preceding paragraph. Produced by the technique is a set of dimensions that (in a certain mathematical sense) may be said to underlie the structure. There is one set of dimensions for the rows ("occupations," in the above example) and one set for the columns ("subjects for a beautiful photograph"). Nonetheless, the two different types of phenomena (those named by the rows and those indexed by the columns) are mapped into a single space, which Bourdieu calls "the space of correspondences." It is because of this duality that correspondence analysis often goes under the name of dual scaling (Levine 1979). Bourdieu often interprets these dimensions with respect to types of capital (e.g., economic capital, cultural capital) and volume of capital (see, e.g., Bourdieu [1979] 1984).

The connection of all this to loglinear modeling is the following. One of the main developments in Goodman’s work in the 1980s and 1990s has been the elaboration of deep formal similarities between models of correspondence analysis and loglinear models, particularly the “association models” that I introduced earlier. Goodman never refers to Bourdieu—after all, Goodman is interested in statistical models, and Bourdieu is interested in social fields—and Goodman’s references to the French methodologists who worked on developing correspondence analysis are few. Goodman is interested in furthering a different conversation (see, e.g., the title of Goodman, 1966), one in which classically different styles of statistical reasoning are "reconciled" and "synthesized." So the French social theorist and the American statistical methodologist are largely unaware of the implications of each man’s work for that of the other. These implications are relevant, nonetheless, for the formal models (Breiger 2000) and for rhetorical development, as I now suggest.

Institutional Relations and Power Relations

Loglinear modeling and correspondence analysis, viewed by many of those who are uninterested in the enterprise and also by most sociology graduate students endeavoring to learn these procedures, are antiseptic activities far removed from contending social institutions and national culture differences. Such is not true of the statisticians and sociological methodologists who produce these models of data analysis.

11. Correspondence analysis ensures that the scale value produced for each row may be written as a weighted sum of the scale values produced for the set of columns (where the weights are proportional to cell frequencies in the table under analysis), and dually for the columns. In this sense the rows, so to speak, "are" the columns, and vice versa.
To begin with correspondence analysis: criticisms of correspondence analysis and related procedures as practiced by French, Dutch, and occasionally Japanese researchers have been legion, especially in America and England. Disputes among statisticians on this subject are so heated as to often result in a rhetoric of national character differences. "These national distinctions are more real than one might think," the eminent British statistician J.C. Gower writes in the pages of Applied Statistics (1989, 273), where he treats correspondence analysis as an instance of "Franco-Dutch statistics" which in confrontation with "Anglo-Saxon statistics" leads to "opportunity for confusion" on core issues of statistical practice: what is and is not a model, the importance or otherwise of graphical methods in defining a model, data description versus statistical testing, and other issues. For his part, Goodman, whose "style or rhetoric" was characterized by Duncan (1984b, xii) as so close to the technical issues of modeling that he "never allows us to relax our responsibility for the fine structure of an analysis in favor of a specious preoccupation with the big picture," nonetheless manages to aver in a leading American sociology journal (Goodman 1987, 535) that

although it is not very relevant, I cannot forbear to include here the following... statement from [a French]... report: "French sociology will not 'follow' Anglo-Saxon sociology... in its mathematisation but will develop its own indigenous methodological research."

From the side of France, J.-P. Benzécri, an early shaper of the development of correspondence analysis in that country, commenting on a paper of Goodman's in a leading American journal of statistical science, and glancing in passing at "the three small tables that served as a grounding point in his article," goes on to proclaim (1991, 1115) that

On both sides of the Atlantic, the same numerical algorithms are applied to data analysis. Bibliographical references cross over the ocean, but the very spirit of correspondence analysis, as we understand it, did not cross over yet.

Here the question of multiplicity—the same numbers are computed robustly on both sides of the Atlantic, but the "spirit" is weak—is inscribed within the differing institutions of national culture. My point is that conversations about loglinear modeling are not only related to communities of practitioners, but that such communities exist within national cultures that themselves provide some of the features to conversations among statisticians that we sometimes associate with communication across ethnic boundaries.
Often the Americans are seen, or are assumed, or are accused, of having the politically privileged position in this international methodological field. The authors of an extremely influential cross-national mobility study centering on European nations (Erikson and Goldthorpe 1992b) endeavor to maintain a contrast between two images, the "mobility" and "achievement" points of view, connoting respectively the topological models they formulate and—on the other hand—the image of a hierarchy of status attainment that they (and Duncan 1979) see as incorporated within the association models used by some American researchers and which Erikson and Goldthorpe (1992b) label "the American dream." In reply to the criticism of Hout and Hauser (1992), Erikson and Goldthorpe (1992b, 298) write that "we think that debate in the field of stratification and mobility research would be . . . made more productive if those pursuing the 'American Dream' were to recognize that . . . it is not in itself a fault for other investigators to have . . . different conceptions of what macro-sociology can and should aim to achieve." And so quantitative social scientists assert linkages between the technicalities of formal models and global ideological hegemony.

Model-Rich Environments

Mailloux (1989, 19) argues with respect to literary analysis that we are in need of rhetorical histories that are interpretive, institutional, and cultural. In this chapter I have argued that very similar sorts of rhetorical histories can help us—whether we are sociological researchers, students, or curious denizens of other disciplines—to understand how formal modeling joins with data analysis and theorizing about our world in an active process that includes and needs to draw in with full consciousness the modeler and analyst. Although rhetorical analysis can of course be used to debunk and dismiss that which we do not like, the more serious scientific task is to deepen our understanding of a world that includes us.12

Literary analysts often speak of subjecting a text to interpretive stress. There is a useful analogy to the way quantitative social scientists such as Otis Dudley Duncan speak (as in Exhibit 2 of this chapter) of "a model-rich environment" as "not necessarily a comfortable one." In this chapter I have argued for a recognition of parallelism in issues of concern to literary analysts and to quantitative social science modelers. I have focused on

problems of multiplicity, reflexivity, duality, and institutional context and power relations.

The implications of this recognition run in both directions. Sociologists could benefit from engaging with questions raised by literary theorists as a means toward reevaluating the discipline’s foundational issues. Pescosolido and Rubin (2000) provide a tentative step in this direction. It is possible to combine interpretive readings of sociologists’ quantitative procedures with more usual methods of textual interpretation in order to understand relations among their contributions, as I have done with reference to writings of James Coleman and Pierre Bourdieu (Breiger 2000). For their part, literary analysts and, in particular, those interested in reception histories, could benefit substantially from the contributions of sociologists to frameworks for the empirical study of the sorts of local cultural practices that are at the center of study of at least some theorists of rhetoric. Note 5 of this chapter suggests some promising leads.

Graduate students in sociology are used to assuming that issues of theory are matters of contention and worth fighting over. When it comes to methods, however, they assume, and we often decline to challenge the assumption, that the quantitative models are somehow neutral, natural, bland (if not boring), and difficult to master (if not entirely irrelevant to theoretical and ideological struggles). Is it any wonder that we sociologists are often accused of poor writing? I urge more self-conscious, more wide-ranging conversation as a means to better writing of quantitative sociology.

BIBLIOGRAPHY


110 Ronald L. Breiger


Zald, Mayer N. 1995. “Progress and Cumulation in the Human Sciences after the Fall.”
Sociological Forum 10: 455–79.

ACKNOWLEDGMENT:

I am grateful for conversations with John Baldwin, Mitchell Duneier, Noah Friedkin, Roger Friedland, John Mohr, Bruce Straits, Thomas P. Wilson, and Raymond Wong during a sabbatical year at the University of California at Santa Barbara, and for continuing discussions with Peter Bearman, Elisabeth Clemens, Gerry Cox, Donald P. Hayes, Keith Hjortshøj, James P. Lantolf, Joel Levine, Calvin Morrill, Philippa Pattison, Alan Sica, David Stark, Henry A. Walker, Linda Waugh, Harrison White, Robin M. Williams, Jr., and Mayer Zald.