Handbook of Methods in
Cultural Anthropology

Handbook of Methods in Cultural Anthropology

H. Russell Bernard
editor

ALTAMIRA PRESS

A Division of Sage Publications, Inc.
Walnut Creek ■ London ■ New Delhi

1998

W. PENN HANDWERKER
STEPHEN P. BORGATTI

Fifteen

Reasoning with Numbers

This chapter explains the reasoning behind numerical methods and illustrates their application to questions we ask when we carry out ordinary ethnographic research. *Numerical methods* encompass a wide variety of interlinked techniques that range from basic mathematical tools (for example, matrix algebra) and simple descriptive statistics (modes, medians, means, percentages, ratios), to visualization and data-reduction techniques (for example, multidimensional scaling and factor analysis), to classical statistical tools (for example, t-tests and regression). *Numerical reasoning* consists of answering important research questions with appropriate numerical methods.[1]

Most of the techniques we illustrate go well beyond the content of introductory statistics courses. Nonetheless, we hope to show that even simple forms of numerical reasoning add important components to ethnographic research, that more sophisticated methods answer questions that may prove crucial in your research, and that explicit reasoning with numbers reveals things you'd otherwise miss. Since this chapter can't substitute for formal study, we direct you to literature that can help you master techniques you find intriguing.

Answering Questions We Ask in the Field

All ethnographic research begins by collecting data from one person. When we go to the next person, we always find something different, as well as something much is much the same. And so it goes. We keep track of similarities, note variability, and keep at it until we no longer find significant cultural or behavioral variability. Then, we construct a story from the inferential generalizations we arrived at about the people we worked with: about their lives and the circumstances in which they have lived; about what those people now think, feel, and do; about who agreed with whom about what and to what extent; and about who is similar to whom and to what extent, and how they differ from others and to what extent.

This considerably simplified account highlights eight activities entailed by ordinary ethnographic fieldwork:

1. We *measure*. Measurement transforms sensory information into intelligible mental constructions. Measurement thus refers to what we do when we make and record observations about what we experience in the field. Much ethnography entails altering the constructions we took into the field in ways that improve their correspondence with those used by the people in our field setting. We also may create new ways of thinking about the world to help us better understand what we experience in the field. In either event, we aspire to an understanding of what we see during fieldwork that corresponds with what our informants tried to teach us and take that understanding a step further. This corresponds with the research design issue of *internal validity* (see Campbell and Stanley 1966).
   - We measure two fundamentally different kinds of data: *cultural* data and *life experience* data. *Cultural* data, as we use the term, are measurements of the systems of mental constructions people use to interpret themselves and the world around them and of the behavior isomorphic with those systems of meaning. *Life experience* data, as we use the term, are measurements of individual characteristics (like age, gender, height, or weight) and events or processes that mark the life experience of particular people (like how many years a person spent in school, whether they grow cash crops, if they grew up in poor or wealthy households, how often they use condoms when having sex, or the degree to which they experienced one or another form of violence as a child). We highlight the operationally important differences between these two kinds of data in later sections of this paper.
   - *Measurement error* constitutes a principal source of invalid research findings. Intensive and prolonged interaction predicated on personal relationships with informants is one check on this source of error. *Triangulation* using a variety of data collection and analysis tools to address any one question or issue helps even more (McNabb 1990).
2. We pay attention to *variables*. Variables consist of phenomena that may vary from one observation to another. Variation in a phenomenon may exhibit qualitatively different assignments of meaning. The variable *kin* may be brother, sister, husband, or wife, for example. *Gender* may be woman, man, girl, or boy.

Variation may exhibit quantity. *Family size* may be small, medium, or large (an *ordinal* form of measurement, which only entails relative rank); or 5, 10, and 20 (a *ratio* form of measurement characterized by a true zero, which allows any two values to be expressed as a meaningful ratio of the other).[2]
3. We aim to identify *what is typical* in one sense or another. For one purpose or another, we may want to know
   - How *most* people conceptualize specific people, events, and processes (measured explicitly by a statistic called the *mode*).
   - The age at which, on average (measured explictly by a statistic called the *mean*), as well as how, children become adults, and people first have sex, marry, give birth, and become grandparents.
   - Important and unimportant people, events, and processes; or people who move into adulthood, first have sex, marry, give birth, and become grandparents early or late. The warrant for judgments about relative placement consists of a middle point (measured explicitly by a statistic called the *median*) that distinguishes important from unimportant, or early from late.
   - The *median* is one of a set of statistics called *quantiles*, which measure relative placement in a distribution. The median, for example, is the 50th percentile. Other quantiles include the 10th percentile (which identifies the location below which lies the earliest or lowest set of cases), the 25th and the 75th percentiles (between which the middle 50% of the cases lie), and the 90th percentile (which identifies the location above which lies the latest or highest set of cases).
4. We evaluate *variation from what is typical*. For one purpose or another, we may want to know
   - How commonly we find a particular way of conceptualizing specific people, events, and processes. We can explicitly measure how commonly something occurs as a simple count (*frequency*) of the number of people who take one point of view or another. The *modal frequency* is the point of view with the highest count. More usefully, we can express this number as a *proportion* (the modal frequency/the total count), a *percentage* (100*a proportion), or a *ratio* (for example, the number of people who express the most common point of view/the number of people who express any other point of view). We might be inclined to generalize a consensus if 80% of the people we talk with express the most common (*modal*) point of view, for example. We probably would infer much cultural diversity if we find many ways of thinking about an issue and only 20% express the most common (*modal*) view. In the former case, the modal view occurs (80%/20%), four times (400%) more often than any alternative; in the latter, it occurs (20%/80%), 25% less often than the alternatives. Great variation suggests the operation of many contingencies that

influence how people conceptualize themselves and the world around them. Little variation suggests the operation of significant constraints.

- How much variation, on average, exists in the age at which children become adults and people first have sex, marry, give birth, and become grandparents (measured explicitly by a statistic called the *standard deviation*). Great variation suggests the operation of many contingencies that influence when and, perhaps, how life course events occur. Little variation suggests the operation of significant constraints bearing on life course events.
- The earliest point at which an event occurs in people's lives (the *minimum*), the latest point that event occurs (the *maximum*), the distance between these points (the *range*), and the relative frequency with which an event occurs between these points (a variable's *frequency distribution*). The distribution of events may be *skewed* over a given range of occurrence—most people marry first at relatively young ages, but some marry first late in life, for example. A distribution may be *symmetrical*, or approximately so, as might be the case for the age at which children undergo one or another puberty rite.
- Events that occur unusually early or late (*outliers*) relative to others.

  Pictures of the distribution of a variable—*stem-and-leaf plots*, *box plots*, or *quantile plots*, for example—make it easy to evaluate these salient characteristics of a variable's distribution.

5. We try to identify the people, events, and processes that go together and those that don't. *Similarity* coefficients, each in different ways, explicitly measure the extent to which two phenomena go together or not. *Dissimilarity* coefficients, each in different ways, explicitly measure the extent to which two phenomena don't go together. Explicit measurements of relationships among even a few people, events, and processes produce hundreds if not thousands of details, which human minds appear particularly unsuited for comprehending. Tools like *factor analysis, cluster analysis, multidimensional scaling, and correspondence analysis*, reduce this complexity and transform these details into pictures, which human minds appear particularly well suited for comprehending.
6. We try to identify who agrees with whom about what and to what extent, which we can measure explicitly with *similarity* coefficients. A set of techniques called *consensus analysis*, which uses a specific form of *factor analysis*, tests for the existence of cultural consensus.
7. We aspire to an understanding of what we see during fieldwork that goes beyond the few people we had a chance to learn from. This aim corresponds with the issue of *external validity* (to whom, if anyone, can we generalize?) and recognizes that, having studied with only a handful of teachers, we generalize to a community of hundreds, perhaps hundreds of thousands of people we never met. This goal signals the difference between using statistics to *describe* characteristics of the people we spoke with (*descriptive* statistics) to using the same statistics to *infer* something about a larger population (*inferential* statistics). Research design and sampling procedures address these questions (see Kish [1965] for sampling of life experience data and Handwerker [1998] for sampling of cultural phenomena).
   - All data collection entails sampling—gathering information from only *some* of an indefinitely large number of people, places, and times. Gathering information from other people, or the same people at different times (the next day) and places (next door), almost always yields different findings even when population characteristics (*parameters*) don't change. We can't tell from looking at differences—whether they appear in our fieldnotes or in survey results—whether they represent something real (people changed, people living in different neighborhoods think and act differently) or not (people haven't changed, and people living in different neighborhoods think and act pretty much the same). *Statistical tests* yield *probabilities* that tell us how often specific differences could occur merely by chance.
   - *Inferential* data analysis consists of *measuring* characteristics of an unbiased ("random") sample of independent cases, computing *statistics*, and *estimating* population parameters. So long as you draw an unbiased ("random") sample of independent cases, any statistic (identified by Latin letters, like b or mean) can be taken as a *point estimate* of a population parameter (identified by Greek letters, like β or μ). The statistic (which you compute) constitutes your single best guess at the parameter (which you can't compute). You can also compute *confidence intervals* for a statistic. Ninety-five percent confidence intervals computed for 100 different samples of the same size from the same population will contain the population parameter 95 times, for example. Classical statistical tests (*chi-squared, t-tests, F-ratios*) yield valid probabilities that tell you how often to expect a specific finding just by random sampling fluctuations—so long as you test an explicit hypothesis.[3]
   - The socially constructed nature of cultural phenomena, however, means that any one person who knows about a particular cultural phenomenon participates with other experts in its construction. This means that a random sample of people does not constitute a random sample of the cultural phenomenon you want to study, since talking with any one person yields data very much like that provided any other (see Handwerker and Wozniak 1997; Handwerker et al. 1997). Although case-dependence like this means that standard kinds of statistical tests won't give us the information we need for valid analysis of cultural data, Weller (1987) has shown that the Spearman-Brown Prophesy Formula can

be applied to informants rather than items, which allows us to measure the reliability and validity of the cultural data we report. Ethnographic findings based on information from small numbers of informants (3–36, depending on the average level of agreement), exhibit exceptional reliability (.90–.99) and validity (.95—~1).

- *Bootstrap* (Efrom and Tibshirani 1991; Mooney and Duval 1993) and *jacknife* (Tukey 1958) procedures create sampling distributions for statistics from the data at hand, and yield valid parameter estimates for life experience data for situations in which the theoretical sampling distribution of a statistic, on which classical statistical tests rely, remains unknown (for example, Handwerker 1988). *Permutation tests* give valid probabilities (Fisher 1935; Edginton 1980) for any set of data. Permutation tests yield exact probabilities by expressing the observed findings (and all more extreme findings) as a ratio of all possible events. ANTHROPAC (Borgatti 1992) offers permutation tests; SYSTAT 7.0 (Wilkinson 1997) has integrated bootstrap and jacknife options into its regression module, and others where appropriate.

8. We may go further than mere description to try to determine why and how the patterns and variability we found might have come into being, why and how they persist over time, and why and under what circumstances they may change. We aim to explain variability in culture and behavior as a function of variability in experience, and search for concrete events and circumstances that shape those experiences. We also understand that what people think and do must reflect not only their individual life history, but broader regional and global histories of people, events, and social interaction into which they were born and in which they grew up. So we try to identify events, circumstances, and processes that provide one set of choices to some people and a different set of choices to others. We ask individuals to identify life experiences that were significant to them, and to help us understand why those experiences were significant.

- Similarity coefficients and statistical tests applied to data like these can tell us what goes with what and how strongly, and thus provide a warrant for believing that a given relationship is *real*, not merely a figment of our imagination (McEwan 1963). Handwerker and Crosbie (1982) showed that the relationship between gender and social dominance (men tend to be dominant over women) is real, for example. Unfortunately, analysis of the relationship between only two variables (*zero-order analysis*) tells us next to nothing in which we can have confidence. Adding another variable to the analysis may produce dramatic changes—a zero-order similarity coefficient may disappear, grow stronger, or change from positive to negative.

- Research that goes beyond cultural description thus raises *internal validity* issues that go beyond the simple problem of *measurement error* mentioned earlier. Warrant for inferring that a relationship is *determinant* as well as *real* (McEwan 1963) requires us to isolate a suspected relationship so we can tell whether or not it exists, and its strength, when we rule out the other internal validity possibilities. *Experimental research designs* (for example, Campbell and Stanley 1966; Cook and Campbell 1979) were created to do just that. Although ethnographic research rarely allows for the use of experimental designs, experimental design principles apply to all research. *Multiple regression* techniques approximate the goals of experimental designs and have many uses in ethnographic research. Handwerker and Crosbie (1982) used a form of multiple regression called *path analysis* to show that, although the relationship between gender and dominance was real, it wasn't determinant: social dominance followed from control over access to resources and, independently, from relative size. Larger people tended to dominate smaller people. Men tend to dominate women only because, generally, men are larger than women. But just as larger women tend to dominate smaller women and larger men tend to dominate smaller men, larger women tend to dominate smaller men.

What Numerical Analysis Does

Numerical methods thus constitute nothing more than explicit tools of data collection and analysis that address core research questions: "Did we get it right?" and "To whom, if anyone, can we validly generalize?" The operationally important difference between one or another numerical method is the question it answers. Each technique answers a different question. Table 1 shows the correspondence between specific numerical methods and specific questions we ask in the course of conducting ordinary ethnographic research. *Numerical reasoning* consists of answering important ethnographic questions with the appropriate numerical method.

Research consists of a search for ways to distinguish mental constructions that consist largely of error from constructions that consist of less. *Validity* consists of a relationship between the definitions of specific mental constructions and specific observations. It follows that we assess data and finding validity with reference to whether or not, or the degree to which, specific mental constructions correspond with specific observations. Being explicit helps immensely when you and others try to find errors in judgment that, once identified, prod you to think about the world in new, more interesting, and, perhaps, more useful ways. Judgment errors come in three main ways. First, you may construct your study phenomena from the wrong components. Second, you may construct arguments with questionable assumptions

TABLE 1
Correspondence Between Selected Numerical Analysis Tools and Questions Anthropologists Try to Answer

NUMERICAL ANALYSIS TOOLS	RESEARCH QUESTIONS: Who agrees with whom about what and to what agree?	What's the agreement about?	What's there to be explained?	Who (what) acts (looks) like whom (what) and to what degree?	What goes with what and to what degree?	Can we see a suspected relationship even after we control for everything else we can think of?
Consensus Analysis	■	■	■			
Multidimensional Scaling (MDS)	■	■	■	■	■	
Quadratic Assignment (QAP)	■	■	■	■	■	
Similarity and dissimilarity coefficients	■	■	■	■	■	
Cluster Analysis	■	■	■	■	■	
Correspondence Analysis	■	■	■	■	■	
Perceptual Mapping/PROFIT Analysis	■	■	■			
Factor Analysis	■	■	■	■	■	
Simple Summary Statistics	■	■	■			
Numerical Transformations						■
Graphical Analysis Tools for Single Variables			■			
Graphical Analysis Tools for 2 or More Variables			■	■	■	■
Survival Analysis			■			■
Ordinary Least Squares (OLS) Multiple Regression			■			■
Logistic Multiple Regression			■			■
Probabilities	■	■	■	■	■	■

or make logical errors. Third, you may mistake what you would like to be with what is. The first source of judgment errors constitutes the focus of debates and arguments over theory. Explicit identification of the assumptions used to construct different theories helps clarify what arguments are all about and uncover logical errors and thus minimize the second source of error. But arguments over logically sound theory are resolvable—albeit only temporarily—only by reference to assessments of the fit between specific constructs and specific observations. Explicit numerical reasoning helps minimize this third source of judgment error and allows us to see things we'd otherwise miss.

Cultural Data: A Look at Families and Kin Relations

Danielle Wozniak's recent research among foster mothers in the United States (Wozniak 1997) addresses a question at the core of cultural anthropology since its inception: What is a "family?" We shall use her data as a running example to illustrate the numerical methods most suited for the analysis of cultural data.

Like most ethnographic research, Wozniak relied on extensive participant-observation over a three-year period, involving day-long periods of observation, two–three hours of taped interviews, and active involvement in the lives of foster mothers. One of the most important themes that emerged in the texts of daily conversations and transcribed interviews involved criteria her informants used to talk about themselves and other women as "mothers" and how other people—husbands, children, social workers, friends, and relatives—used or didn't use these criteria in equivalent ways. She felt these women telling her that how they assessed the relative importance of motherhood criteria was at least as important—probably more important—than knowing which criteria they used to make sense of themselves and their activities as mothers.

Wozniak asked women to evaluate the importance of nine criteria her informants had identified as ones used, not used, or used differently, when they thought about themselves and other women as mothers, carrying out mothering activities. These criteria included that a child: (1) be born to a woman; (2) be born to a family member; (3) contribute work or income; (4) be loved by a woman; (5) be taken care of by a woman; (6) live with a woman for a long time; (7) be adopted; (8) be thought of by the woman as being the child's mother; and (9) that the child think of the woman as her or his mother. These questions thus asked about conventionally recognized legal, biological, and instrumental criteria as well as self-defined affective criteria that foster mothers had mentioned. To avoid having the order of presentation influence women's responses, Wozniak randomized the order in which women were asked these questions.

Sixty-nine foster mothers—Black, white, rich, poor, old, and young—rated each criterion on a four-point scale from "not at all important" to "very important."

Scales like this can be treated either as an ordinal sequence of quantities or qualitatively different assignments of meaning. In most subsequent analyses, we treat the points on these scales as qualitatively different assignments of meaning.

Matrices and Matrix Operations

All numerical analysis begins with the construction of a data matrix. Typically, matrix rows correspond to individual informants (*cases*) and the columns express what they tell us or what we see (*variables* and their *attributes* or *values*). Each column consists of a vector of numbers or characters, and a matrix consists of a collection of vectors. The following matrix, for example, shows a portion of Wozniak's data on the race of her foster mother informants and their rating of various affective motherhood criteria. Wozniak's data comprise a *cross-sectional profile matrix*. Her data are cross-sectional in that they were collected for a specific point in time. The data also include specific information for each individual informant and thus give us a profile of each. If Wozniak had collected data on the same variables for two or three sequential periods in time (for example, Day 1, Day 10; or 1985, 1995), her matrix would consist of *panel* data (for example, see Markus 1979). If Wozniak collected data on the same variables for an increasing number of sequential time periods (for example, Day 1, Day 2, Day 3 . . . Day k; or 1950, 1955, 1960 . . . 1995), her matrix would consist of *time-series* data.

RACE$	AFFEC1	AFFEC2	AFFEC3	AFFEC4
W	4	4	4	4
B	4	4	4	4
B	4	4	4	4
W	4	4	3	3
W	4	4	4	2
B	4	4	4	3
B	4	4	4	4
B	4	4	4	2
W	4	4	4	4
W	3	4	4	3

Numerical analysis consists of operations carried out on matrix rows or columns (*vectors*) or both (for example, Namboodiri 1984). For example, summing each column of numbers and dividing by the number of cases provides a convenient summary (the average, or *mean*) of the importance of each motherhood criterion. Counting the number of Ws and Bs tells us the number of Euro and African American women in the sample. Dividing one of these sums by the total number of cases tells us the proportion of Euro and African American women Wozniak interviewed; multiplying that number by 100 converts them to percentages (47.8% of her informants identified themselves as African American). We might want to recode the affective criteria to create *binary* (sometimes called *dummy*) variables with only two values, "1" and "0." If we assign a "1" to any value greater than 2 and "0" for all other values, the sum for each row would tell us the number of criteria that each woman felt important. The average of 1s and 0s for any one motherhood criterion would tell us the proportion of women who felt it important.

When we apply other descriptive statistics mentioned earlier to Wozniak's data, we learn, among other things, that one person was a newly licensed foster mother who had not yet fostered a child (the minimum) and another had fostered around 250 children (the maximum); and that the women who answered her questions ranged in age from 28–78 and averaged 44 years of age, give or take (the standard deviation) about 9 years. Answers to questions like these help us evaluate the extent to which Wozniak's sample differs from all currently licensed foster mothers, but they don't tell us anything interesting about the central question: What criteria do you apply to determine whether or not, or the extent to which, you are a mother? We need to apply other matrix operations.

Figure 1, for example, juxtaposes stem-and-leaf and box-and-whisker plots of four variables, one representing biological criteria of motherhood (gave birth to the child), one representing instrumental criteria (a child's contribution of work or income), one representing legal criteria (adoption), and one representing affective criteria (giving care to the child) (see, for example, Tukey 1977; Cleveland and McGill 1985; and SYSTAT's graphics module documentation). Stem-and-leaf plots identify the minimum and maximum scores (1 and 4), the median, and the upper and lower hinges (the 25th and 75th percentiles). The plot stem identifies the values (1, 2, 3, 4); one leaf appears next to each value for each informant who reported that assessment. Box plots tell us the location of the middle 50% of cases with a box. These boxes are notched to show an interval that contains the value of the middle case (median) for the entire population 95% of the time (a *confidence interval*). Whiskers extend to the lowest and the highest 25% of cases. Unusually high or low cases (*outliers*), as on the last plot, appear as dots. Examination of these plots reveals, rather dramatically, that Wozniak's informants expressed much variation in their assessments of the importance of biological, instrumental, and legal criteria of motherhood, and little about the importance of an affective criterion. Indeed, so many women rated the affective motherhood criteria as "very important" that the median and both hinge cases are buried at a single point, and the box-plot shows no box. All other ratings (3, 2, or 1) constitute outliers.

Assessing Similarities and Differences

This finding raises two questions: (1) Do these women see these criteria as the same kinds of things, or differently?; and (2) Which women agree with which others

```
Stem and Leaf Plot of variable:    BIO1, N = 69
    Minimum:      1.000
    Lower hinge:    1.000
    Median:     2.000
    Upper hinge:    3.000
    Maximum:    4.000

      1 H 0000000000000000000000000
      2 M 0000000000000000000000
      3 H 00000000
      4   00000000000000

Stem and Leaf Plot of variable:    INSTRU1, N = 69
    Minimum:      1.000
    Lower hinge:    1.000
    Median:     2.000
    Upper hinge:    3.000
    Maximum:    4.000

      1 H 00000000000000000000000000
      2 M 00000000000000000000000
      3 H 00000000000000000
      4   000

Stem and Leaf Plot of variable:    LEGAL, N = 69
    Minimum:      1.000
    Lower hinge:    2.000
    Median:     3.000
    Upper hinge:    4.000
    Maximum:    4.000

      1   00000000000000
      2 H 000000000000000000000
      3 M 000000000
      4 H 0000000000000000000000000

Stem and Leaf Plot of variable:    AFFEC3, N = 69
    Minimum:      1.000
    Lower hinge:    4.000
    Median:     4.000
    Upper hinge:    4.000
    Maximum:    4.000

      1   0000
      2   0
      3   0000000000
      4 M 000000000000000000000000000000000000000000000000000000
```

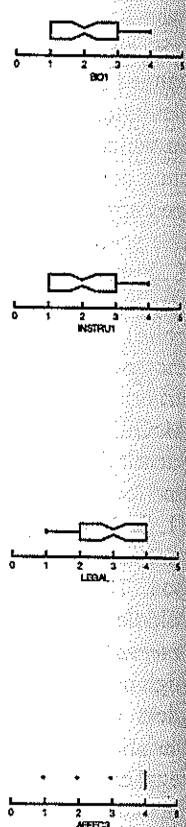

Figure 1. Stem-and-leaf and box plots of foster mother's assessments of the importance of different mothering criteria.

about the nature and relative importance of these criteria? Matrix operations that tell us how much any two cases or variables exhibit similarity or difference yield a wide variety of similarity and dissimilarity coefficients (for example, Sneath and Sokal

1973). Common *similarity coefficients* include Pearson's r (*the* correlation coefficient), r^2, and, for binary data, Driver's G, the positive matching coefficient, the simple matching coefficient, and Jaccard's coefficient. Common *dissimilarity coefficients* include Euclidean distances, city-block distances, positive nonmatching coefficients, and simple nonmatching coefficients.

For ratio data,[4] Pearson's r, which tells us the extent to which variation in one phenomenon accompanies variation in another, varies between ±1.00. Positive coefficients tell you that one variable goes up when the other goes up; negative coefficients tell you that one variable goes up when the other goes down; values close to ±0.00 tell you that what one variable does has nothing to do with what the other variable does. The coefficient of determination, r^2, tells the proportion of variation shared by two phenomena. Euclidean distance coefficients express the square root of the summed squared distances. The Euclidean distance is zero when the variables are identical and gets larger as the values of the two variables get more and more different. City-block distance coefficients just constitute the sum of the absolute distances (like the number of miles between cities, or the length of different city blocks).

Similarity and dissimilarity coefficients for binary data (present-absent) come from cross-tabulation tables like the following:

		X: 1	X: 0	
Y	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	N

Y and X represent variables, like whether or not a person assesses "giving birth" or "adoption" as important criteria for "motherhood." The letters *a,b,c,* and *d* represent the number of informants who fall into particular cells of the table. For example, the value of *a* would tell us the number of foster mothers who thought that both "giving birth" and "adoption" constituted important motherhood criteria. The value of *d* would tell us the number of foster mothers who thought that neither "giving birth" nor "adoption" constituted important motherhood criteria. Sums across rows *(a+b, c+d)* and down columns *(a+c,b+d)* yield marginal frequencies, which sum to N, the total number of cases.

Positive matching coefficients tell you the number of times that X and Y share the presence of something *(a)* as a proportion of the total (N). Simple matching coefficients tell you the number of times that X and Y share both the presence and absence of something *(a+d)* as a proportion of the total (N). The dissimilarity positive and simple nonmatching coefficients constitute mirror images of their similarity cousins. By contrast, Driver's G and Jaccard's coefficients ignore phenomena X and Y do not share (the frequency of cell *d*). Matching coefficients, under some circumstances, make some people, households, societies, and organisms look similar largely because they share so little. To avoid this problem, Jaccard's coefficient tells you the number of times that X and Y share the presence of something (*a*) as a proportion of all of the times X and Y share something (*a*+*b*+*c*). Driver modified Pearson's correlation coefficient for binary data by eliminating the value of *d*:

$$\text{Driver's G} = a/\ (\text{sqr}(a+b)*\text{sqr}(a+c))$$

Analysis of similarities or dissimilarities among sets of variables begins with the construction of matrices made up of similarity or dissimilarity coefficients, like the following Pearson correlation matrix for the motherhood criteria Wozniak studied.

The unities (1.000) across the diagonal just tell us the obvious—variables covary perfectly with themselves. The interesting coefficients tell us the degree to which women assessed the different motherhood criteria in identical ways. The correlation coefficient of .585 tells us that Wozniak's informants assessed giving birth (BIO1) and children contributing work and money (INSTRU1) in reasonably similar ways. They assessed children contributing work and money (INSTRU1) and giving care to children (AFFEC3) in essentially unrelated ways (the correlation of .041 is very close to zero).

	BIO1	INSTRU1	LEGAL	AFFEC3
BIO1	1.000			
INSTRU1	0.585	1.000		
LEGAL	0.445	0.361	1.000	
AFFEC3	0.246	0.041	0.308	1.000

Number of observations: 69

It's hard to say much more with confidence about the relationships among all variables. Eye-balling the coefficients suggests that women tended to assess the importance of biological, instrumental, and legal criteria in similar ways, in ways related little if at all to their assessment of affective criteria. But the correlation of .308 between the legal criterion and the affective criterion clouds this judgment. To make clearer evaluations of how women made their assessments, we need to analyze a matrix showing the interrelationships of all nine variables simultaneously. Trying to make judgments about 36 similarity coefficients at once just intensifies the information overload we've stumbled into (Miller 1956). Moreover, sets of similarity coefficients may exhibit relationships due solely to chance (for example, see Handwerker 1991).

Data-reduction techniques that make this information load manageable include factor analysis (see Rummel 1970; Harman 1976; see also Driver and Schuessler

1957), cluster analysis (see Sokal and Sneath 1963; Sneath and Sokal 1973; Aldenderfer and Blashfield 1984), and multidimensional scaling (see Kruskal and Wish 1978). These techniques also help us distinguish matrices with *structure* (relationships among variables that don't exist solely by chance) from matrices with none (relationships that exist solely by chance).

Factor Analysis

Let's turn to the question of whether women assessed the importance of biological, instrumental, and affective criteria in similar ways, in ways related little if at all to their assessment of affective criteria. If they did, that suggests that women experience and respond to biological, instrumental, and legal criteria in ways fundamentally different from the way in which women experience and respond to affective dimensions of motherhood. Indeed, distinctions like this may exist only on the edge of consciousness. Women might not have the words to describe such differences, which nonetheless may significantly influence how they experience and respond to the world, and shape their family relationships in fundamental ways.

Factor analysis constructs a small set of variables (*factors* or *principal components*) from additive combinations of existing similarities among variables or cases.[5] Each factor infers the existence of one or more underlying, unmeasured variables that might explain the observed pattern of similarities. Handwerker (1997), for instance, used factor analysis to show that people who live dramatically different lives—descendants of former plantation slaves who make a living in tourist economies in the West Indies and people who continue to hunt, gather, and herd in the Alaskan and Siberian Arctic—agree about components that comprise unitary phenomena legitimately called "violence" and "affection." In a factor analysis, the first principal component extracted accounts for the maximum amount of variance in a matrix. The second factor or principal component accounts for the maximum amount of the remaining variance. The subsequent principal components account for the maximum amount of the variance left by previously extracted factors.

Factor analysis output includes correlations between each factor and either your cases or variables, called *loadings*. The square of a loading tells you how much variability a case or variable shares with the unobserved variable represented by each factor. The sum of squared loadings for a factor (its *eigenvalue* or *latent root*) tells you how much variation *all* your cases or variables share with a factor. A factor's eigenvalue divided by the sum of eigenvalues for all factors expresses the factor's explanatory power as a proportion.

Look at the eigenvalues from a principal components analysis of the similarity matrix for all motherhood criteria (SYSTAT output):

1	2	3	4	5	6	7	8	9
3.908	1.714	0.936	0.764	0.567	0.425	0.292	0.256	0.138

Note that the eigenvalues drop precipitously after the first and level off after the second, where they fall below 1. Judgments about how many real factors a matrix may contain should reflect close analysis of the eigenvalues. Matrices that contain real underlying structure exhibit a very sharp drop in the size of eigenvalues and a clearly identifiable point at which eigenvalues level off (shown in *scree plots* [after the debris, called scree, that accumulates at the base of cliffs; see Figure 2]). Contrast the first scree plot with the second, which comes from eigenvalues generated by an identical principal components analysis of similarities among variables related only by chance.

The kinship data yielded only two of nine eigenvalues over 1.0 whereas the second set of data yielded five eigenvalues over 1.00, five under 1.00, and no perceptible scree. The differences between the two distinguish matrices containing real factors (like Wozniak's data on mothering) from matrices with no underlying structure (random relationships). Just by chance, sample items may show great similarity. Factor analysis of matrices like this will find some eigenvalues over 1.0 and some high loadings just by chance. But matrices that contain no real factors will generate random distributions of eigenvalues and loadings.

Latent Roots (Eigenvalues)

1	2	3	4	5	6	7	8	9	10
1.633	1.299	1.204	1.144	1.088	0.986	0.891	0.634	0.618	0.503

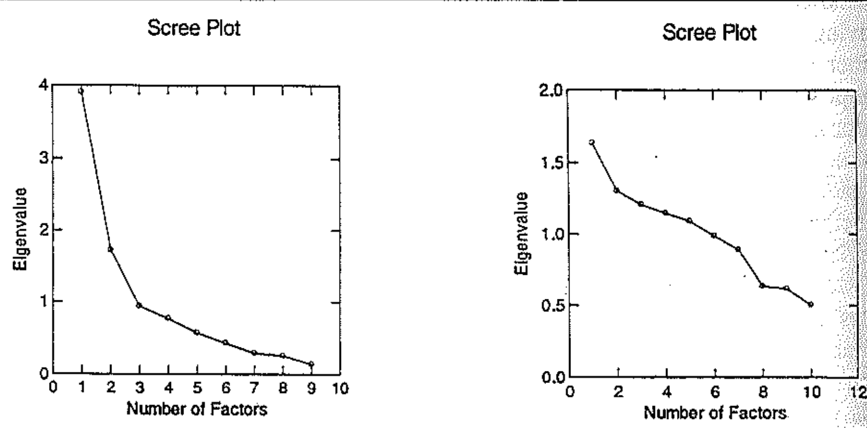

Figure 2. Scree plot of eigenvalues from principal components analysis of (left) foster mothers' assessments of the similarity among mothering criteria and (right) similarities among variables related only by chance.

Further analysis of Wozniak's kinship data focused on the first two factors identified in the scree plot in Figure 2. An operation called *rotation* clarifies the differences between factors. Rotation using the Varimax criterion maximizes the differences between the factors. The rotation process redistributes the proportion of variation each factor explains, but the total remains the same. Two factors account for just over 62% of the variability in these women's assessments of the relative importance of nine criteria bearing on mothering and mothering activities:

Rotated Loading Matrix (VARIMAX)

	1	2
BIO1	0.863	0.066
BIO2	0.851	0.125
LEGAL	0.709	0.332
INSTRU1	0.708	0.033
INSTRU2	0.544	0.436
AFFEC2	0.093	0.835
AFFEC1	0.018	0.798
AFFEC3	0.156	0.765
AFFEC4	0.321	0.690

Variance Explained by Components before Rotation

1	2
3.908	1.714

Percent of Total Variance Explained

1	2
43.425	19.048

"Variance" Explained by Rotated Components

1	2
2.907	2.716

Percent of Total Variance Explained

1	2
32.295	30.177

Note that this output also sorted the variables by loading size, for loadings equal to or larger than .5 (loadings of .5 or larger mean that a factor and a variable share at least 25% of the observed variance). BIO1, BIO2, LEGAL, INSTRU1, and INSTRU2 load highly on one factor; all AFFEC variables load highly on the other. These findings suggest that Wozniak's informants responded to formal biological, legal, and instrumental motherhood criteria in one way and responded to personal affective criteria in another. Cluster analysis gives similar results.

Cluster Analysis

Cluster analysis assigns cases or variables to groups based on relative similarity or dissimilarity. There are a large variety of clustering methods, based on different algorithms and different criteria for constructing groups.

In a study of the class structure of Mexican villages, for example, Allison Bingham (1995) used a *k-means cluster analysis* to show that the historical *ejidatario-avecendados*/landed-landless class structure had been completely reconfigured. K-means cluster methods split cases (households, in this instance) into hypothesized numbers of groups maximizing the differences between groups and minimizing the differences within groups. The *ejidatario-avecendados* (two-class) distinction held when she looked only at legal land ownership, but fell apart as she added other pertinent variables. What began historically as a two-class structure had acquired multiple levels. The wealthiest community members owned stores and/or drew on income from migrant family members, and turned liquidity into de facto land ownership, acquiring land from *ejidatarios* who had no other sources of credit for farm operations.

Probably the most commonly used clustering methods, however, use Johnson's hierarchical clustering technique, which itself contains several variations. The basic variants apply *single-link* or *complete-link* criteria to matrices of similarity or dissimilarity coefficients. Both variants begin by assigning each item being clustered (cases or variables) into a cluster by itself, yielding N clusters (where N is the number of items). Then it determines which pair of clusters is the most similar (or the least dissimilar) and combines the two, resulting in N–1 clusters.

This process continues until only one cluster remains, which contains all the items. The difference between single-link and complete-link methods consists of the criteria applied to determine similarity between existing pairs of clusters. Single linkage procedures match up cases or variables based on the similarity between a candidate member and any *one* existing member.

Complete linkage procedures match up cases or variables based on the similarity between a candidate member and *all* existing members. Figure 3 shows a complete linkage cluster analysis of women's assessment of motherhood criteria. The results tell us that women assessed affective criteria one way and biological, instrumental, and legal criteria in another. Contrast the complete linkage clusters of motherhood criteria with complete link clusters of random relationships.

Multidimensional Scaling (MDS)

Multidimensional scaling (MDS) transforms a matrix of similarities or differences into a map. MDS transforms a matrix of distances among cities in the United States (city-block dissimilarities) into map coordinates of the United States, for example. MDS coordinates thus express strong similarities (small dissimilarities) as spatial closeness and weak similarities (great dissimilarities) as spatial distance.

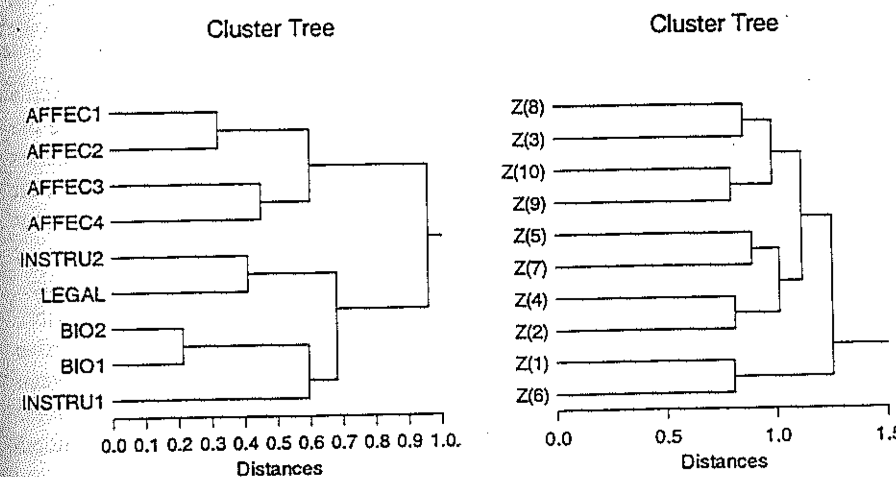

Figure 3. Cluster trees of foster mothers' assessments (left) of similaries among mothering criteria and similarities (right) among variables related only by chance.

MDS transforms similarity and dissimilarity coefficients into coordinates for two, or for several dimensions. In a study of dance among Northern Ute and the predominantly Mormon Anglos living in adjacent communities, for example, Stephanie Reynolds (1990) used a three-dimensional MDS analysis to show that, whereas Anglos differentiated the meanings of dance, the meaningfulness of dance, its purpose and function, and clearly separate it from their history and way of life, Utes equated the meanings of dance, the meaningfulness of dance, its purpose and function, with each other and with Ute history and way of life. In a study of Alaskans affected by the Exxon Valdez oil spill, Joseph Jorgensen (1995) used three-dimensional MDS to show that natives configured their subsistence economy in ways qualitatively different from nonnatives, despite undertaking many of the same activities, living in the same communities. Where natives experienced spiritual dimensions to the environmental damage and responded communally, nonnatives, sensitive to commoditized properties of the environment, responded as individuals.

Like the Cartesian coordinates that mapmakers use, the axes of an MDS picture are arbitrary and uninterpretable. The only thing that matters is the relative placement of cases or variables on the MDS map. Two-dimensional coordinates map a plane; three-dimensional coordinates map volume; larger numbers of dimensions probably can't be interpreted. A value called *stress* measures the extent to which mapping distorts the distance between pairs (of cases or variables). Stress varies between 0 and 1. A stress value of 0.0 means that no distortion took place. Stress values below .10 constitute evidence of an excellent fit between the matrix and the mapped coordinates; stress values up to around .15 exhibit at least reasonable levels of fit; stress values over .20 call for additional dimensions.

Figure 4 shows an MDS plot of the similarity coefficients among women's assessments of various motherhood criteria (stress = .075). The MDS plot, like the factor analysis and the cluster analysis, suggests the existence of two broad classes of motherhood criteria that women may experience and respond to in very different ways: personal and formal. It suggests, further (like the cluster analysis), that different kinds of affective criteria may exist that reflect important differences about how foster mothers experience being mothers and carry out mothering activities. Text (from conversations and formal interviews) hinted at the former distinction, which these analyses confirm. The possible difference between kinds of affective motherhood criteria remained hidden. So did a possible equivalence between adoption, the biological child of a relative, and living together for a long time, which only the MDS map suggests. Contrast the MDS map of the motherhood criteria with the mapping of randomly generated variables.

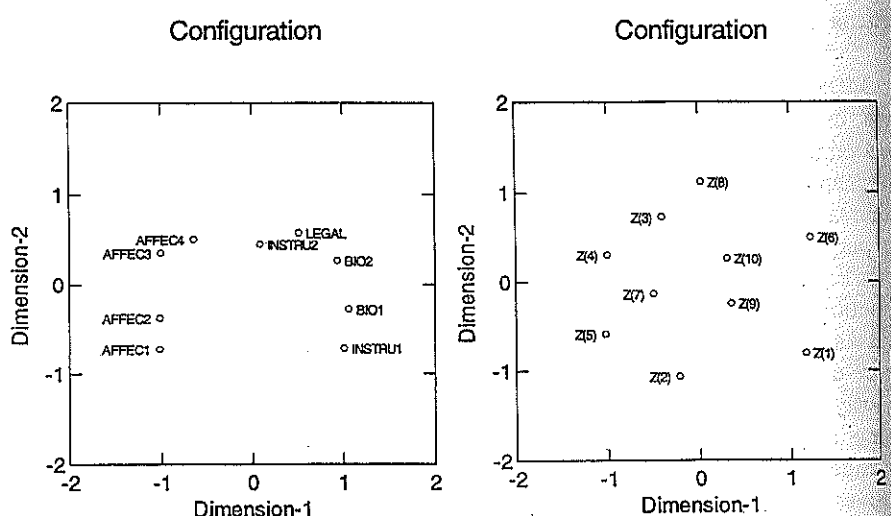

Figure 4. MDS maps of (left) foster mothers' assessments of similaries among mothering criteria and (right) similarities among variables related only by chance.

Consensus Analysis

Consensus analysis (for example, Romney et al. 1986; Weller and Romney 1988; see Boster 1981, 1987, 1988) answers what may be the single most important question of ethnography: Who agrees with whom about what and to what degree? Consensus analysis thus answers the second question raised by examination of the stem-and-leaf and box plots of Wozniak's data: Which foster mothers agree with whom about which motherhood criteria and to what degree?

Consensus analysis is a singularly valuable analytical tool that contributes directly to the ongoing reorientation of theory and research in ways that transform ethnography into ethnology by a focus on cultural variability between individuals rather than between reified and essentialized groups (see Bidney 1944; Barnett 1953; Wallace 1961; Murdock 1971; Pelto and Pelto 1975; Wolf 1982; Keesing 1994). For a given domain of meaning and behavior, variation in the data we collect in the field may reflect three conditions:

1. random variation around a single consensus about the domain or some aspect of it, which may be weak or strong;
2. subpopulation differences in the strength of agreement with the single consensus, or two or more systematically different sets of meanings and behavior, which may differ little or be polar opposites; or
3. no consensus about the domain or some aspect of it.

The first condition identifies the absence of *cultural boundaries* (after Keesing 1994) within a given population and, thus, the absence of significant cultural variability. The second condition identifies the existence of cultural boundaries within a given population and, thus, the presence of significant subpopulation variability. The third condition signals the presence of such widespread disagreement within a population that the data can't be distinguished from random data.

Historically, we have used assumptions, not evidence, to equate cultural boundaries with social identities like Nuer, Navaho, and African American, and have dismissed, overlooked or downplayed both cultural variation among people who use the same social identity and cultural consensus among people who use different social identities. Consensus analysis procedures and diagnostics allow us to test explicitly for the presence or absence of cultural boundaries and to identify disagreements bearing on specific domains of meaning and behavior.

Consensus analysis thus promotes an important new rigor in cultural description and explanation. In a study of American culture, for example, Kempton et al. (1995) found consensus about environmental values that was integrated with other core American values, including religion and parental responsibility, and which was shared by people as diverse as members of Earth First! and laid-off sawmill workers. They found little antienvironmental sentiment and no coherent set of values that contrasted with or was an alternative to the coherent and widely shared proenvironmental values.

William Dressler used consensus analysis to construct regionally and historically specific measures of poverty (see Dressler 1996, 1997; Dressler et al. 1996) based not on the conventional and narrow biological conception of need, but on one more germane to understanding meaning and behavior—relative deprivation in lived experiences (for example, Aberle 1966). The resulting measure of *cultural consonance* encompasses the lived experience of poverty with its multiple dimensions—income, education, age, marital status, occupation, and social support and integration—and thus offers a new, explicit tool for understanding the sources, patterning, and implications of disadvantage in human communities.

Dressler's work documents clear health implications. Similarly, efficient sampling strategies for cultural data focus on the selection of knowledgeable informants whose lives encompass experiential variability that may influence the phenomena studied, track levels of agreement, and expand sample sizes and sampling criteria consistent with levels of agreement and identified cultural boundaries (Handwerker and Wozniak 1997). Consensus analysis combined with explicit sampling frames addresses the problems of internal and external validity and yields findings generalizable to a population defined by an explicitly identified set of life experiences (Handwerker 1998).

Consensus analysis procedures came from a theory that defines culture as knowledge and that distinguishes variation due to the use of different configurations of knowledge (usually called subcultural variation) from variation due to differences in how much individuals know about a given cultural domain. For example, Boster (1981) noticed that the Aguaruna women who agreed most often with others on questions about a set of manioc plants were also the ones who knew the most about it. Similarly, on a multiple-choice examination in which the teacher defines the culturally correct answer key, students who know most of the answers tend to agree a lot with each other because they agree closely with the common answer key. At the same time, students who know very little of the material give answers that agree little with anyone.

Romney et al. (1986) showed that one can use patterns of agreement to reliably infer how much informants know about a domain when three conditions hold: (1) There is a single right answer to every question; (2) Conditional on the answer key, people's answers to questions are independent of each other (that is, students are not cheating and answering one question does not affect one's answer to another); and (3) Questions are drawn at random from one domain of knowledge (for example, the test doesn't mix questions about tennis with questions about anthropology). Given these conditions, we don't need the answer key to work out how much of the material each student knows. A factor analysis of a chance-adjusted person-by-person agreement matrix estimates a person's knowledge score based on the correlation between their responses and everyone else's. Once we estimate how much each person knows, we can infer the right answers by the application of *Bayesian* statistical methods.

Consensus analysis procedures thus lend themselves particularly well to the analysis of cultural domains for which there exist unique culturally correct answers and variation in knowledge about the domain. Examples include the correct way to give birth, hold girls' or boys' puberty rites, or marry—or any form of *evaluation research*. Evaluation research tests the efficacy of interventions designed to induce specific forms of cultural change. Example interventions include bicycle safety programs designed to increase the chances that a child will use a helmet while riding, training of health care providers designed to improve their ability to carry out accurate physical examinations, and lesson-delivery that requires fifth-grade students to participate in the construction of mathematics problems designed to improve mathematics problem-solving skills.

Interventions assume the existence of, and teach, "correct" answers—riding with, not without helmets; the correct way to conduct physical examinations; the skills to achieve correct solutions to specific mathematics problems. Evaluation research assesses the degree to which trainees exhibit "correct" answers. People subject to intervention activities constitute "cases;" people not subject to intervention activities constitute "controls." One way to state the evaluation research hypothesis is that cases exhibit greater cultural similarities with the "correct" answers than do controls. Thus, a direct test of the evaluation research hypothesis comes from a consensus analysis carried out on a matrix that includes the intervention-defined answers coded as 1 (correct), the answers given by cases coded as 1 (if correct) or 0 (if not), and the answers given by controls coded as 1 (if correct) or 0 (if not).

The consensus procedure in ANTHROPAC generates four key sets of output data: (1) eigenvalues of each factor, the percentage of variability in informant responses that each factor accounts for, and the ratio of the largest eigenvalue to the next largest; (2) factor loadings (ANTHROPAC's KNOWLEDGE scores), which identify the correlation between the responses of individual informants and the first factor, and the average level of agreement among all informants; (3) the similarities among the responses of your informants (ANTHROPAC's AGREE data file), on which the factor analysis was performed; and (4) an answer key. The first set provides provisional information with which to judge whether or not your informants express a consensus about the issue you're studying. A large proportion of the variance among individuals explained by a single factor the eigenvalue of which is three or more times larger than the next largest—a dramatic scree fall—constitutes evidence (subject to diagnostic analysis) of the existence of a single answer key around which all variation in responses is draped. The last three sets provide critical diagnostic information. Negative factor loadings (KNOWLEDGE scores), for example, signal the possible presence of two or more systematically different sets of meanings within the population. A run of particularly low factor loadings (KNOWLEDGE scores) warns you of significant subpopulation cultural differences.

In the evaluation research example, for example, evidence that the intervention worked and that cases acquired higher levels of competence (wearing helmets, physical exam procedures, mathematics) than did controls consists of a factor loading (KNOWLEDGE score) for the intervention-defined answers approximating 1.00, high factor loadings (KNOWLEDGE scores over .5) for cases, and low factor loadings (KNOWLEDGE scores under .5) for controls. We could count the number of intervention-defined correct answers provided by cases and controls and use a permutation test to find out how often we could find those differences just by chance. However, consensus analysis procedures allow us to look for intervention-induced alternatives or opposing sets of agreements, as well as subpopulation differences in the degree of agreement with intervention-defined correct answers. Consensus analysis thus adds a rich complement to the more straightforward technique.

In the general case, a large number of negative or low factor loadings tell you to disaggregate your data (for example, see Boster and Johnson 1989) and possibly return to your informants to search out the sources of difference. Is this a domain for which there exist acknowledged cultural experts and the low or negative knowledge scores reflect the ignorance of particular informants? Do there exist two or more systematically different sets of meanings and behavior? Are there domain items about which people show no agreement?

Although the theory that led to consensus analysis conceptualized culture as a mental phenomenon, consensus analysis procedures apply equally well to the behavior isomorphic with a given body of knowledge (observational data on how women give birth, how puberty rites are carried out, how people marry—and whether people act differently after an intervention). Data appropriate for consensus analysis might come from text collected by various forms of informal or semi-structured interviews (so long as all informants responded to the same questions), from structured observations, or from structured interview formats like sentence frames, pile sorts, triads tests, or rating scales.

Moreover, since the factor analysis infers the existence of a coherent (albeit otherwise unobserved) agreement about a domain or some aspect of it, consensus analysis procedures also lend themselves to the analysis of cultural domains where the metaphor of an answer key fits awkwardly. In many if not most cultural domains, everyone is a cultural expert. All women are experts about being women; all men are experts about being men. Everyone living in the United States is an expert on life in the United States, and all Americans are experts on American environmental values, for example. Lack of agreement doesn't imply that one person knows more than another. In these circumstances, consensus analysis procedures allow us to infer a coherent, shared agreement about meanings or behavior.

A first eigenvalue massively larger than the second signals a clear consensus and ANTHROPAC's answer key tells us what the agreement is all about. But factor

loadings tell us the strength of agreements, not how much someone knows about the domain. Low or negative factor loadings signal cultural differences arising from life experiences that differ in significant ways. Diagnostic analysis can tell us whether low or negative factor loadings mean a weak consensus with generally low levels of agreement, subpopulation differences in the level of agreement, or the existence of alternative or opposed subpopulation agreements about the domain or some domain items.

Consensus analyses of Wozniak's data, treated as qualitative assignments of meaning (multiple choice data), yielded the following:

	Eigenvalues			
Factor	Value	%	Cum %	Ratio
1:	25.451	61.3	61.3	2.830
2:	8.994	21.6	82.9	1.266
3:	7.103	17.1	100.0	

Note: It would be better if the first factor accounted for more than three times the variance of the second.

Although the first factor accounts for 61.3% of the variability in informant responses, the second accounts for 21.6%. A consensus may exist among some women about some criteria, but these findings suggest the possibility of subpopulation cultural differences. Subpopulations might differ in the degree to which they concur with the overall consensus. There may be two or more coherent sets of agreements. A subpopulation might strongly agree among themselves and sharply disagree with the majority. A subpopulation might agree with some of the overall consensus and strongly agree among themselves about an alternative set of answers to some domain items. To know, we need to disaggregate the data. Maybe Black foster mothers thought about what it takes to be a mother differently than did white foster mothers. Maybe cultural differences reflected how long a woman had fostered, her age, whether her biological children were grown, or whether she was poor.

Subsequent diagnostic analysis, guided by informant interview texts, revealed that none of these speculations bore out. Indeed, they revealed another dimension to the distinction between personal and formal criteria of motherhood examined earlier. *All* women—Black or white, rich or poor, old or young—agreed on the importance of the self-defined affective criteria of motherhood. Consensus analysis of respondent evaluations of the importance of affective criteria yielded a first eigenvalue 7.4 times larger than the second, which accounted for 83.5% of the variability in informant evaluations; the average level of agreement among informants was .76. By contrast, there existed no agreement about the conventional biological, legal, and instrumental criteria of "motherhood." For these latter criteria, the ratio of the first to the next largest eigenvalue was only 1.3, the first factor accounted for only 42.9% of variability in informant evaluations, and the average level of agreement among informants was only .01.

The similarity matrices among informant cases analyzed by consensus procedures may exhibit one of at least four different patterns. First, a very high degree of consensus among informants would appear as a tight centralized cluster of cases with increasing degrees of random scatter as one moved away from the answers that constituted the common agreement (although you might not see the fried-egg pattern if high average levels of agreement place everyone in the core). Second, a weaker consensus might appear as a more uniformly distributed, but still clearly clustered, scatter of cases. Third, the absence of consensus would appear as a random scatter of cases. Finally, the existence of two or more distinct consensuses—as would exist, for example, if Black and white foster mothers responded to the affective dimensions of mothering in different ways—might appear as two or more foci in the MDS map. Figure 5 shows the MDS map of agreements concerning the importance of affective criteria for judging oneself a "mother" (stress = .0517) with the MDS map of agreements concerning the importance of biological, legal, and instrumental criteria (stress = .1237). The first exhibits the very strong consensus identified by the factor analysis (the plot points contain a small amount of random variation—jittering—otherwise the cluster in the center would have appeared as a single point, it was packed so densely). The second exhibits random scatter.

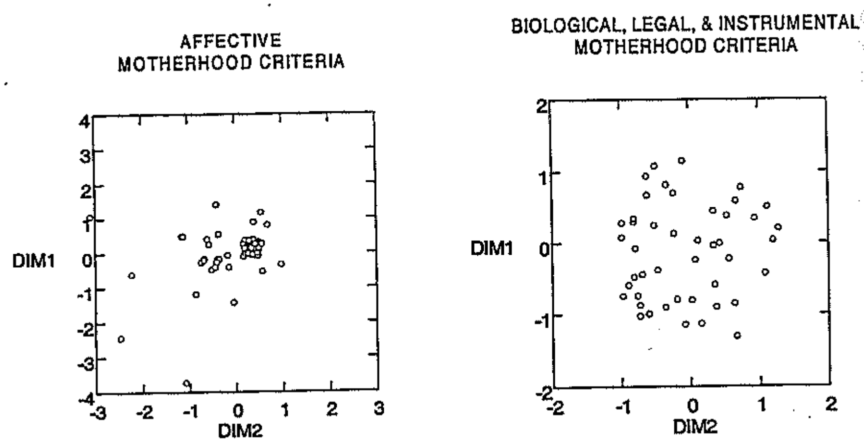

Figure 5. MDS maps of informant agreement about mothering criteria.

In short, through their responses to structured questions about the relative importance of different mothering criteria, Wozniak's informants told her that personal,

affective dimensions of mothering guided and rationalized their mothering activities in important ways. Black or white, rich or poor, old or young, foster mothers experienced formal dimensions of mothering that receive wide social recognition—giving birth to a child, receiving help from a child, adopting a child—as very different from, albeit in almost idiosyncratic ways, the core affective dimensions of mothering and family relationships.

Triangulation

Numerical analysis integrated into the actual fieldwork process builds method triangulation to assess data validity and reliability (see McNabb 1990). By helping you see things you'd otherwise miss, it also identifies questions that lead to a greater depth of understanding. MDS maps and other plots are particularly handy output to take to informants, individually or in focus groups. For example, findings from consensus analysis of Wozniak's informants' assessments of family members and their permanency revealed a highly puzzling finding: Foster mothers thought of foster sons as "very" permanent family members but thought of foster daughters as only "somewhat" permanent family members. She presented this and other findings to focus groups of foster mothers and asked if they thought the findings were true and, if so, why. Group discussions revealed that foster mothers experienced a family dynamic common to biological mothers: Once foster daughters reached their teens, foster mothers and daughters often began to experience conflicts over the daughter's sexual and reproductive behavior. Characteristically, these conflicts escalated and culminated with the pregnant daughter leaving or being told she must leave.

Foster mothers and daughters subsequently had little if any contact for several years. Once daughters established their own families, however, they usually reestablished relationships with their foster mothers. One woman's foster son might well be the father of another woman's foster daughter's child. But foster sons were "very" permanent family members: sons brought home the baby and the girlfriend and foster mothers incorporated both into their families.

Life Experience Data: Modeling Relationships Among Lives, History, and Region

Numerical reasoning thus provides explicit validation for the inferential generalizations about cultural phenomena that we arrive at during the course of ordinary ethnographic research. Ordinary Least Squares (OLS) regression (for example, Schroeder et al. 1986; Gujarati 1995) provides explicit tests for speculations about why and how the patterns and variability we found might have come into being, why and how they persist over time, and why and under what circumstances they may change.

Conceptualizing human beings in ways that coherently incorporate the implications of being born at a specific time and place, the effects of maturation, and the experience of specific historical events and processes at specific ages, constitutes a major theoretical challenge. Historical and regional explanation appears essential. However, historical and regional explanations usually create linkages between temporal sequences of events and processes with ad hoc or post hoc claims that people act for specific intentions, out of specific dispositions, or for specific reasons.

Long experience with (and analysis of) reasons, intentions or motivations found in text provided by informants living or long dead tells us that this information *cannot* decide these issues (see Brown 1963; Hammel 1990). Indeed, this information may mislead us far more than they provide us insight. People lie. People forget. People rationalize what they do. People do things for many reasons. People aren't aware of all or even most of the influences on what they do. People may be completely *unaware* of the most important influences on what they do, particularly when those influences are historical, macrolevel phenomena that can't be perceived clearly in the minutiae of day-to-day living. People misjudge the relative importance of their reasons, intentions, or motivations. Shweder (1977) and many others have documented how easy people find it to tell what is *like* what and how hard people find it to tell what *goes* with what.

Moreover, attempts to explain specific behavior from specific reasons, intentions, or motivations necessarily produce covert tautologies because, if we "explain" a specific behavior by reference to a *different* reason, intention, or motivation, we'd contradict ourselves. Reasons, intentions, motivations, and other internal mental states consitute part of our *descriptions* of cultural phenomena—which call for explanation.

One way to avoid this dilemma is to link theoretical or empirical generalizations to initial conditions established by specific (albeit ungeneralizable) temporal sequences of events and processes. Historical events and processes in specific regions establish initial sets of conditions that people may experience and to which they may respond in predictable ways.

For example, explanations for childbearing during adolescencerange from speculations about the influence of (1) the operation of evolved neural architectures designed to optimize reproductive success (for example, see Tooby and Cosmides 1992), to (2) social circumstances, to (3) gender and intergenerational inequalities bearing on means by which women gain access to resources. An important formulation for the first appeals to girls' early sensitive learning-period responses to father-absent rearing, or to maternal messages about father-absent rearing (Draper and Harpending 1982, 1988). The most popular example of the second appeals to the frustrations, stresses, and strains of poverty. Childbearing, however, might reflect social relations.

On the principle that one must eat before one can reproduce effectively, our central nervous system should have an evolved neural architecture designed to optimize access to resources, and humans gain access to resources through social relations. If childbearing yields the best resource access, as it did in the West Indies when women were dependent on men for income and women used childbearing to legitimize claims of support from men during their youth and to equalize gender inequalities by middle age, women should bear lots of children. If good jobs requiring high levels of educational attainment yield the best resource access, as occurred recently in the West Indies with structural change in regional economies that equalized gender inequalities, women should go far in school and have few if any children. Early sensitive period learning may bear on optimizing resource access, as well as reproductive success. Women who experience exploitative violence (sexual abuse) as children may act in ways that correspond more closely with optimum resource access than their unexploited peers (for example, see Liem et al. 1992). Sexually abused girls thus might have more children than their peers, when childbearing yields the best resource access, and more education than their peers, when education yields the best resource access.

Multiple Regression Models

We indicated earlier that multiple regression methods approximate the goals of experimental designs and help us isolate relationships so we can tell whether or not they exist, and their strength, even after we control for other variables. Answers to "why" questions consist of a claim that one phenomenon is linked to another. Multiple regression is a way of thinking about these claims as general functions. General functions of the form $Y = f(X_k)$ constitute claims about the existence of a set of rules that allow us to translate values of *independent* variables (X_k)—explanatory variables like father-absent rearing, poverty, and inequalities—into values of one or more *dependent* variables (Y)—like adolescent childbearing. A simple additive (*linear*) model to test the possibilities outlined earlier looks like the following:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

where each X represents one of the three explanatory variables and the Greek epsilon (ϵ) constitutes explicit recognition that we will make the best predictions possible, but we anticipate prediction error. Alternatively, we could write the model as: adolescent childbearing = f(father-absent rearing, poverty, and inequalities), and write the equation with variable names:

$$\text{ADOLESCENT CHILDBEARING} = \beta_0 + \beta_1\text{FATHER-ABSENT REARING} + \beta_2\text{POVERTY} + \beta_3\text{INEQUALITIES} + \epsilon$$

The Greek betas (β) represent the relationships that may exist between Y and each explanatory variable. Assume that we've measured the independent variables as simply present (1) or absent (0).

> If there exist no father-absent rearing, no poverty, and no inequalities (all Xs equal 0), β_0 (the *constant*) shows the predicted number of children born during adolescence (ADOLESCENT CHILDBEARING = $\beta_0 + \beta_1*0 + \beta_2*0 + \beta_3*0 = \beta_0$).
>
> If father-absent rearing exists but there exist neither poverty nor inequalities, $\beta_0+\beta_1$ shows the predicted number of children born during adolescence (ADOLESCENT CHILDBEARING = $\beta_0 + \beta_1*1 + \beta_2*0 + \beta_3*0 = \beta_0+\beta_1$).
>
> If father-absent rearing, poverty, and inequalities all exist, $\beta_0+\beta_1+\beta_2+\beta_3$ shows the predicted number of children born during adolescence (ADOLESCENT CHILDBEARING = $\beta_0 + \beta_1*1 + \beta_2*1 + \beta_3*1 = \beta_0+\beta_1+\beta_2+\beta_3$).

The *regression coefficients* (β_1, β_2, β_3) thus tell you how much adolescent childbearing changes with any change in the independent variables; regression coefficients may be negative or positive numbers, so they also tell you whether the number of children born goes up or down.

Introducing Complexities from the Intersection of Lives, History, and Region

Simple linear models fit an amazing number of real world situations. Simple modifications transform them in ways suitable to many real world nonlinearities and situational, historical, and regional contingencies. For example, in poor homes adolescent childbearing may rise proportionally (may double, for example), not by fixed amounts (by one child, for example). The following model, which transforms the dependent variable units into logarithms, tests that possibility.

Ln(ADOLESCENT CHILDBEARING) = $\beta_0+\beta_1$FATHER-ABSENT REARING+β_2POVERTY +β_3INEQUALITIES+ϵ

Similarly, father-absent rearing, poverty, and inequalities may have nothing to do with adolescent childbearing—unless they all occur together (an *interaction* effect). The following equation tests this possibility:

ADOLESCENT CHILDBEARING = $\beta_0+\beta_1$FATHER-ABSENT REARING+β_2POVERTY +β_3INEQUALITIES+β_4(FATHER-ABSENT REARING*POVERTY *INEQUALITIES)+ϵ

If father-absent rearing, poverty, and inequalities (for *direct* effects of the variables) exert no influence on the number of children girls bear before they turn 20 unless they occur together, you couldn't distinguish the size of the first three regression coefficients from 0.00, and the final model would appear like:

ADOLESCENT CHILDBEARING = $\beta_0 + \beta_4$(FATHER-ABSENT REARING * POVERTY * INEQUALITIES) + ϵ

Regression also facilitates thinking of what people experience with *systems* of equations and, thus, modeling *pertinent* intersections of lives, history, and region. The regional and historical contingencies summarized earlier suggest the following (truncated and simplified) system of equations bearing on adolescent childbearing (see Handwerker 1989, 1993a, 1993b):

> Childbearing leveled gender inequalities by a woman's middle age, for women subject to gender inequalities early in life. So, the extent to which women conceptualized childbearing as an investment activity rather than a consumption activity (their moral economy of childbearing) = f(job opportunities for women).
>
> Women found education useful only to the extent that it gave them access to good jobs. So, educational attainment (by age 20) = f(moral economy of childbearing in the absence of job opportunities for women).
>
> Given that childbearing levels gender inequalities by middle age, for women subjected to gender inequalities early in life: Childbearing during adolescence (to age 20) = f(moral economy of childbearing*length of time spent in one or another sexual union, for women with no biological constraints on fecundity, educational attainment, and the absence of job opportunities for women who experienced childhood sexual abuse).

This system of equations implies that adolescent birthrates will decrease and educational attainment will increase as job opportunities for women grow, and that women who experience adolescent childbearing will invest heavily in childbearing early in life, but only in the absence of job opportunities. In the presence of good job opportunities, these women will invest heavily in education.

We can evaluate systems of equations like this with *path analysis,* which uses *standardized* regression coefficients. Standardized regression coefficients allow ready comparison between different sets of data. For specific sets of data, however, unstandardized regression coefficients yield far more interesting findings because they provide a concrete picture of what happens in people's lives.

OLS Regression

Ordinary Least Squares (OLS) regression operations try to account for variability in a set of observations (some girls bear more children during adolescence than others). For any one equation, multiple regression operations estimate parameters (β_k) with statistics (b_k). OLS regression makes these estimates using the same criteria applied to the calculation of means. The result is a set of *regression coef-* *ficients* (b_k) which estimate values of the dependent variable (Y) in ways which guarantee that the sum of prediction errors (the absolute differences between the value of each case and the model estimate for the case, called *residuals*) equals 0.0. Each coefficient tells you the average change in Y for each change in one inde- pendent variable (X), after controlling for the effects of other independent variables.

The following SYSTAT output shows a test of the third equation listed earlier using data collected in Antigua:

DEP VAR: Number of Live Births by Age 20
N: 97
ADJUSTED SQUARED MULTIPLE R: .828
STANDARD ERROR OF ESTIMATE: 0.249

VARIABLE	COEFFICIENT	T	P2 TAIL
CONSTANT	0.752	3.317	0.001
Trajectory set by moral economy*Years in legal unions	0.688	7.634	0.000
Trajectory set by moral economy*Years in commonlaw unions	0.117	6.641	0.000
Trajectory set by moral economy*Years in visiting unions	0.051	6.070	0.000
Educational attainment by age 20	−0.064	−3.482	0.001
No job opportunities for girls sexually abused <16	0.602	3.103	0.003

The regression coefficients (the column of numbers labeled COEFFICIENT) tell us rates of childbearing during adolescence. The model's CONSTANT tells us that these Antiguan girls bore an average of about one (.752) child before age 20, irres- pective of effects from independent variables. The remaining coefficients describe the effects of each independent variable. The coefficient of .688, for example, describes the birth trajectory for the interaction between the degree to which a girl looked on childbearing as a consumption or an investment activity (their moral economy score, which ranges from 0 to 4) and how long they spent in a legal marriage. For each year these Antiguan girls were legally married, for example, they bore an average of .688 children for each increase in the degree to which they thought of childbearing as a consumption or an investment activity. Childbearing rates fall off in commonlaw unions (.117) and visiting unions (.051). In concrete terms, this means that Antiguan girls who thought of childbearing only as a consumption activity (moral economy score of 0) bore an average of (.117*0), or 0 children, each year they spent in a commonlaw union. Antiguan girls who thought of childbearing as an extremely important investment activity (moral economy score of 4) bore (.117*4), or .468 children each year they spent in a commonlaw union—one child every two years, on average. Model estimates of the number of children born to any one girl before age 20 would include the constant (.752) and the effects of other independent variables.

For example, the coefficient for educational attainment by age 20 tells us that, for each year of schooling, the estimated number of children born by age 20 fell by .064 children. Thus, the model estimates that an Antiguan girl in a commonlaw union for five years who thought of childbearing as an extreme investment activity (moral economy score of 4) who had only 10 years of schooling (incompleted secondary education) bore (.752) + (.117*5*4 = 2.34) − (.064*10 = .64), or two (2.452)—maybe three—children by age 20. By contrast, if the girl thought of childbearing as a consumption activity (moral economy score of 1) and had completed 15 years of schooling, the model estimates that she would bear no children by age 20: (.752) + (.117*5*1 = .585) − (.064*15 = .96) = .377.

Sexually abused girls who grew up when the Antiguan economy offered women few good job opportunities, however, bore an additional .602 children before they were 20. Thus, for girls

- in a commonlaw union for five years,
- with only ten years of schooling,
- who thought of childbearing as an extremely important investment activity, and who *also*
- experienced childhood sexual abuse *and* grew up when few good job oppor- tunities existed for women,

the model estimates (2.452 + .602 = 3.054), or about three children born during adolescence.

How can you tell if the model really works? Regression analysis divides the variation in the dependent variable (childbearing during adolescence), measured as the sum-of-squares of the dependent variable, into two components: (1) the residual sum-of-squares (5.641), unaccounted for by the equation; and (2) the regres- sion sum-of-squares (29.018), which tells how much variability the model explains. These quantities appear in the ANALYSIS OF VARIANCE portion of the regres- sion output, which follows this paragraph. A regression sum-of-squares that is large relative to the residual sum-of-squares suggests that something in the model works very well in predicting the dependent variable.

Expressing the regression sum-of-squares as a proportion of the total (regression + residual) gives R^2, which will grow each time you add an independent variable even if the variable doesn't have much effect. The adjusted R^2 of .828 tells you that the model accounts for about 83% of the variation in adolescent childbearing among these Antiguan women, even after controlling for the number of variables in the model. Removing the squaring from the residual sum-of-squares, once you divide it by the number of cases (minus the number of variables in the model), gives you a coefficient equivalent to a standard deviation, called the *standard error of the estimate*. The standard error of the estimate (.249) tells you how much error you make, on average, each time you estimate the dependent variable. This model's average error rate is about one quarter of a child. These numbers are remarkable for data collected from individuals, where R^2s of .25 or lower may signal important findings.

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	29.018	5	5.804	93.617	0.000
RESIDUAL	5.641	91	0.062		

What appear as extremely strong relationships can find their way into our data merely by random sampling fluctuations (chance), however. The R^2 of .828 estimates a parameter ρ^2 (rho-squared). The F-ratio output (93.617) gives a probability, which tells you how often to expect the observed R^2 or any larger just by chance. The listed probability for the model as a whole (.000) tells you that we could get an R^2 of .828 or larger less than once in a thousand just by chance. This warrants the inference that something about the model works better than chance in estimating variation in adolescent childbearing among these Antiguan girls. *T*-tests tell you how often to expect the observed regression coefficients just by chance. The high *t*-test statistics correspond with low *t*-test probabilities, which tell you that each independent variable predicts adolescent childbearing in ways most unlikely to occur merely by chance.

OLS Regression Diagnostics

Inferences such as these depend on random sampling and an ideal world of data collection, in which you:

- include all pertinent and no irrelevant or extraneous variables,
- correctly specify the functional form of all supposed relationships between dependent and independent variables,
- measure all variables perfectly,
- predict dependent variables in such a way that the errors we make exhibit a normal distribution,
- select cases in such a way that independent variables exhibit no correlation, and
- findings of the model solution generate no correlated residuals.

Of course, the real world and the ideal world don't often correspond very closely. Sometimes, they match badly. Much data collection and analysis involves trying to determine which of these conditions are met acceptably and which are not, and correcting particularly bad matches between the ideal world and the real one. For example, the first assumption tells you to make your final evaluation of model variables by including only those variables for which you find evidence (low probabilities) that they belong. If you believe that you've met the other assumptions, do an appropriate post hoc test with all variables that otherwise appear extraneous to test for otherwise hidden additional effects. The following SYSTAT output comes from a post hoc test for significant additional effects on adolescent childbearing of growing up (1) in poverty; (2) with a stepfather; (3) in a stable nuclear family household; or (4) only with a mother; (5 and 6) the experience of father-absent rearing (measured directly by the presence or absence of the biological father during an hypothesized early sensitive learning period, and as father-absent rearing messages received from the mother; and (7) age (as a control for historical change). The high probability (.557) tells us that the effects of these variables could be found by chance nearly 60 times in 100. This finding warrants the inference that these variables either have no effect on adolescent childbearing, or effects so small we can't detect them.

TEST OF HYPOTHESIS

SOURCE	SS	DF	MS	F	P
HYPOTHESIS	0.369	7	0.053	0.841	0.557
ERROR	5.272	84	0.063		

The second assumption (correct functional form) tells you to assess possible nonlinearities. If relationships between variables exhibit proportional changes rather than fixed unit changes, logarithm transformations will produce a model that fits the data better. The third assumption (perfect measurement) cannot be fulfilled in the real world. But this assumption bears more on the precision of model estimates than on their accuracy. Random measurement error makes it harder to see real-world relationships, and measurement imperfections in the dependent variable exhibit minor effects compared with measurement imperfections for independent variables.

Violations of the fourth assumption (normally distributed residuals) rarely invalidate findings. However, violations of the fifth assumption (no correlation among independent variables) pose major dilemmas, and violations of the sixth (no correlation among model estimates) are deadly. These conditions mean that the F-ratio and t-tests don't yield valid probabilities. Sophisticated diagnostic techniques now allow us to evaluate whether or not our data violate these assumptions, and equally sophisticated techniques give us means to avoid the analytical confusion such violations create (for example, Gujarati 1995).

Explicit randomization in true experimental designs meets the fifth assumption, since it guarantees that relationships among independent variables exist only by chance. All other data violate the assumption that there are no relationships among independent variables. For all practical purposes, all data sets contain *multicollinearity*. Nonetheless, multicollinearity doesn't always invalidate probabilities. The following conditions signal the presence of enough multicollinearity to distort model estimates:

- *Condition indices* greater than 30 (see Belsley et al. 1980).
- Including a new variable increases rather than decreases the standard error of the estimate.
- A new variable exhibits a high probability and *also* raises the probability of a variable already in the model.
- The model exhibits a low F-ratio probability but exhibits no low t-statistic probabilities.
- Your software tells you that it can't solve for the coefficients.

Factor analysis offers one solution to multicollinearity problems. If the troublesome collinear variables load highly on a single factor, you can measure all simultaneously with *factor scores.* Factor scores measure the contribution of each variable to the underlying factor and thus integrate all variables into a single score. If you suspect that significant multicollinearity still gives you invalid probabilities, identify a promising model, enter additional (control) variables one-by-one, and report the t-tests.

Violations of the sixth assumption, that prediction errors associated with any one case remain uncorrelated with errors made for other cases, appear as patterned relationships among residuals called *heteroskedasticity*. For example, errors in estimation (*residuals)* may be small when you estimate low values of the dependent variable, and grow increasingly larger as you estimate increasingly high values of the dependent variable. Estimation errors may exhibit high similarity for people who live close to each other and decreasing similarity for people who live increasingly far away, a pattern now called *spatial autocorrelation.* Similarly, estimation errors may exhibit high similarity for cases that occur close to each other in time and decreasing similarity for cases increasingly far away, a pattern now called *temporal autocorrelation.* Models that make the dependent variable in one equation an independent variable in the other frequently produce OLS solutions in which the errors for dependent variable estimates correlate with the same variable used as an independent variable.

Finally, all cultural data necessarily entail a combination of both spatial and temporal autocorrelation. This observation dates from the first time an anthropologist used numerical methods—Edward B. Tylor's (1889) attempt to explain the origins and distribution of kin avoidances with a hodgepodge of psychological, functional, and evolutionary hypotheses. One of the originators of the numerical tools that we use today, Sir Francis Galton, asked Tylor how he knew his cases were independent, since the similarities he tried to explain by reference to individual needs and social functions might merely reflect social interaction mediated either by a common history or geographical propinquity. Tylor had no answer.

The second time an anthropologist used numerical methods, Franz Boas (1894) addressed this problem with a cluster analysis of regional and historical social relationships bearing on the distribution of myths and stories found among tribes of

the Northwest Coast of North America, the adjacent plateau, and three widely scattered communities. Boas found that neighboring communities shared more than distant communities (for example, Northwest Coast tribes were more similar to each other than to tribes on the plateau, but Northwest Coast and plateau tribes shared more than either did with the interior Athapaskan, Ponca, and Micmac) and that people who may share a common ancestral community (who spoke languages from the same language family) shared more than people who spoke languages from different language families. Boas thus demonstrated case dependence in cultural data collected in and taken to characterize different communities.

Methods for diagnosing and correcting spatial and temporal autocorrelation may be fruitfully applied when the unit of analysis corresponds to an explicitly identifiable regional or historical unit. *Measurement unit transforms* (for example, logarithm transformations), *weighted least squares*, and *differencing* procedures may eliminate the correlated errors associated with cross-sectional heteroskedasticity and many forms of temporal and spatial autocorrelation (see Odland 1988). *Two-Stage Least Squares* (2SLS) procedures eliminate the correlated errors produced by sets of equations in which independent variables in one equation appear as dependent variables in the other. More recent research, which extends Boas's findings to cultural data collected from and taken to characterize different individuals (Handwerker and Wozniak 1997),[6] suggests that the complexity of spatial and temporal autocorrelation among individuals defeats corrective techniques like these. Cultural data require the application of the Spearman-Brown Prophesy formula to assess reliability and validity or bootstrap, jacknife, or permutation test probabilities, depending on the question you need answered.

Findings that survive exhaustive diagnostic tests warrant substantive interpretation. For example, the system of three equations listed earlier belongs to a more extensive set which tests hypotheses bearing on sexual precocity (when adolescent girls first have sex), sexual mobility (how many sexual partners they have during adolescence), their moral economy score, their educational attainment, and their adolescent childbearing. Solutions for each subsequent equation yield estimates for variables in later equations. Simulations based on these estimates help clarify the implications of the findings. If we assume that all models hold up under rigorous diagnostic scrutiny and we create simulations which fall into the range of values observed for all variables, we find that sexually abused girls invested in childbearing more than their peers if they grew up when the Antiguan economy offered few good job opportunities for women: Sexually abused girls (who averaged only 10 years of schooling, and 6 if they grew up in poverty) bore an average of two children and girls who weren't (who averaged 12 years of schooling, and 11 if they grew up in poverty) bore an average of one. We also find that sexually abused girls invested in education more than their peers, if they grew up after the Antiguan economy offered good job opportunities for women: Sexually abused girls who averaged zero children by age 20 attained an average of 15 years of schooling and girls who weren't who also averaged zero children by age 20 attained an average of 13 (12 if they grew up in poverty).

Conclusion

Our numerical analysis tool kit has grown prodigiously over the last decade or so, as a quick look through the contents of good software programs like SYSTAT and ANTHROPAC shows. At the beginning of this chapter, Table 1 summarized the correspondence between selected numerical methods and the principal questions we ask in the course of ordinary ethnographic research. But many more tools exist. *Andrew's Fourier Plots* and *Icon Analyses* provide graphical alternatives for diagnosing cultural consensus, cultural differences, and disagreements, for example. *Correspondence analysis* provides a graphical means to look at which informants tell us what (Greenacre 1984). *Property-Fitting (PROFIT)* analysis (in ANTHROPAC) and *Perceptual Mapping Analysis* (in SYSTAT) provide ways to assess the meanings that guide informant judgments about similarities or dissimilarities (for example, Carroll 1972). Decisions bearing on whether a test designed for one population (for example, a test for disability among middle-class Euro American children) can be validly applied to another (for example, children living in Puerto Rico) require a distinctive set of *test item analytical procedures* (for example, Gannotti 1998).

Evolutionary and developmental change—in which phenomenal properties at one point in time (t_1) are contingent on phenomenal properties at an earlier point in time (t_0)—can best be evaluated with *Guttman scaling* analyis (for example, Goodenough 1944, 1963; Carneiro 1962). *Linear programming* operations (for example, Chvatal 1983, Feiring 1986) provide tests for optimization hypotheses. When we change the analytical question from *How many* children does a girl bear on average before she turns 20? to What's *the likelihood* that a girl will bear a child before she turns 20? OLS regression yields impossible estimates (under 0 and over 1) and correlated residuals. When we analyze binary or qualitative dependent variables, *logistic regression* (Hosmer and Lemeshow 1989) substitutes for OLS regression to provide us modeling capabilities sensitive to situational, historical, and regional contingencies. *Time series* (see, for example, Gujarati 1995 or any good econometrics text) and *event-history* analyses (for example, Yamaguchi 1991) require specialized numerical methods. We need a distinctive set of numerical methods even to *conceptualize* the population processes and demographic characteristics so central to an understanding of human-environmental relationships and much cultural evolution and social change (for example, Handwerker 1990). In short, numerical methods apply to an indefinitely large number of questions.

Numerical reasoning often constitutes the only available means of answering important questions. For example, *Quadradic Assignment Procedures* (ANTHROPAC's *QAP*; see Hubert and Schultz 1976) give us the capability to test for relationships between multidimensional phenomena measured with a large number of variables on two populations or subpopulations. Allison Bingham (1998) measured activity patterns with more than 20 movement variables (for example, washing clothes, taking care of children, sitting out in the evening) and used QAP to test for differences in movement patterns between people with malaria and those without. James Boster (for example, see Boster and Johnson 1989, Kempton et al. 1995) has used QAP analyses to test for subpopulation differences in consensus agreement patterns.

Similarly, time-series graphics allow us to set cultural and life experience phenomena in context with historically and regionally specific processes. Handwerker used a *Fourier blob icon analysis* of a macroeconomic time-series for Antigua, W.I. to pinpoint the dates of structural change in the national economy. This made it possible to identify precisely which informants could provide first-hand accounts of the early years, how their world changed, and how they experienced those changes; which informants could provide first-hand accounts of the transition years, the conflicts that arose and how they came to be resolved (if they were); and which informants could provide first-hand accounts of life once the structural changes clarified themselves, and how they see themselves differ from people in the older cohorts. By pinpointing the dates of change, moreover, this analysis also provided a foundation for construction of a microlevel measure of "employment opportunities for women" from the aggregated data on macrolevel changes in the Antiguan economy (Handwerker 1993b). This variable effectively identified powerful influences on women's view of childbearing and their reproductive behavior.

Unaided, our senses tell us about surface phenomena—the net effects of chance and complexly textured multidimensional influences on what we think, feel, and do that operate all the time. Numerical methods help us identify chance fluctuations in what we see, detect the subtle configurations of meaning that remain unspoken and, perhaps, unrealized, by our informants, and both identify and assess the relative importance of the multidimensional historical, regional, situational, and individual influences on meaning and behavior. Explicit reasoning with numbers thus lets us see things in and about people's lives we'd otherwise miss and answer questions in ways that help us pinpoint those errors in judgment that make us quintessentially human.

NOTES

We thank Robert Bee and several anonymous referees for helping us avoid unnecessary embarrassment by pointing out important ambiguities, errors, and other lapses in good judgment in earlier drafts. We take sole responsibility for what you see now.

1. As a first-year graduate student at Berkeley studying under A. L. Kroeber (Driver and Kroeber 1932), Harold Driver created a very effective similarity coefficient (Driver's G, which Ellegard [1959] worked out independently, along with its sampling distribution, 30 years later). Driver carried out the first explicit reliability assessment of cultural data (1938) and has written excellent reviews of the development and use of numerical methods in anthropology up to the last quarter of the twentieth century (1962, 1965, 1970). Jorgensen edited a volume (1974) that assesses Driver's own considerable accomplishments.

2. You will find ratio scales called interval scales. However, you will rarely come across true interval scales like the Farenheit and Celsius temperature scales. Interval scales like these do not contain a true zero, since 0 measures a degree of temperature (molecular movement), not the absence of temperature (as does 0 in the Kelvin temperature scale).

3. Computing a set of statistical tests to look for low probabilities produces invalid probabilities. Just by chance, you'll find low probabilities—5% of the time, you'll find probabilities equal or less than .05. You'll find it useful to see this yourself. Follow the procedures outlined in Handwerker (1991).

It's also a mistake to look just at probabilities. A large enough sample will show you that almost everything exhibits a nonchance relationship with everything else. Similarity coefficients for that sample will also show you that nearly all of these relationships are very, very weak and, so, inconsequential. Although anthropologists rarely use samples greater than 100 cases and almost never use samples greater than 500, you still need to focus on the *strength* of relationships, for relationships that don't appear due to sampling fluctuations.

4. Pearson's r computed for ordinal data usually is called Spearman's Rho; Pearson's r computed for binary data usually is called Phi.

5. We describe a *principal components* analysis. Classical forms of factor analysis, to which purists restrict the factor analysis name, reverse this procedure by looking for additive factors that account for observed similarities.

6. Indeed, Handwerker's appreciation of numerical methods owes much to both history and geographical propinquity. He received his degree from the University of Oregon, where he was heavily influenced by two lines of descent from Franz Boas, the first American anthropologist to publish a numerical analysis of cultural data. One line of descent runs through Alfred Kroeber and Harold Driver to his teacher, Joseph G. Jorgensen. The other runs through Melville Herskovits to his teacher, Vernon R. Dorjahn. The seeds of his understanding of cultural phenomena come from another, which runs through Alfred Kroeber to his teacher, Homer Barnett. We might learn much about cultural phenomena from a study of the social interaction influenced by descent and geography within our own discipline.

REFERENCES

Aberle, David F. 1966. *The Peyote Religion among the Navaho*. Chicago: Aldine.

Aldenderfer, M. S., and R. K. Blashfield. 1984. *Cluster Analysis*. Quantitative Applications in the Social Sciences, Vol. 44. Beverly Hills, CA: Sage Publications.

Barnett, Homer G. 1953. *Innovation*. New York: McGraw-Hill.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics*. New York: John Wiley.

Bidney, David. 1944. The Concept of Culture and Some Cultural Fallacies. *American Anthropologist 46*:30–44.

Bingham, Allison. 1995. Household Economic Differentiation among Small Coffee Producers: A Case-Study from Soconusco Region, Chiapas, Mexico. Paper delivered at the 1995 meeting of the American Anthropological Association, Atlanta.

Bingham, Allison. 1998. The Malaria Ecosystem in Southern Mexico. Ph.D. diss., Storrs: University of Connecticut.

Boas, Franz. 1894. *Indianische Sagen von der Nord-Pacifischen Kuste Amerikas*. Berlin: Asher.

Borgatti, S. P. 1992. ANTHROPAC 4.0. Columbia, SC: Analytic Technologies.

Boster, James S. 1981. How the Exceptions Prove the Rule. Ph.D. diss., Berkeley: University of California.

Boster, James S. 1987. Introduction. *American Behavioral Scientist 31*:150–162.

Boster, James S. 1988. "Requiem for the Omniscient Informant": There's Life in the Old Girl Yet. In *Directions in Cognitive Anthropology*. Janet Dougherty, ed. Pp. 177–197. Urbana: University of Illinois Press.

Boster, James S., and Jeffrey C. Johnson. 1989. Form or Function: A Comparison of Expert and Novice Judgments of Similarity among Fish. *American Anthropologist 91*:866–889.

Brown, Robert. 1963. *Explanation in Social Science*. Chicago: Aldine.

Campbell, D. T., and J. C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Carneiro, R. L. 1962. Scale Analysis as an Instrument for the Study of Cultural Evolution. *Southwestern Journal of Anthropology 18*:149–169.

Carroll, J. D. 1972. Individual Differences in Multidimensional Scaling. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. 1. R. N. Shepard, A. K. Romney, and S. B. Nerlove, eds. Pp. 105–155. New York: Seminar Press.

Chvatal, Vasek. 1983. *Linear Programming*. San Francisco: W. H. Freeman.

Cleveland, William S., and Robert McGill. 1985. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science 229*:828–833.

Cook, T. D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Draper P., and H. Harpending. 1982. Father Absence and Reproductive Strategy: An Evolutionary Perspective. *Journal of Anthropological Research 38*:255–273.

Draper P., and H. Harpending. 1988. A Sociobiologial Perspective on the Development of Human Reproductive Strategies. In *Sociobiological Perspectives on Human Development.* K. B. MacDonald, ed. Pp. 340–372. New York: Springer-Verlag.

Dressler, William W. 1996. Culture and Blood Pressure: Using Consensus Analysis to Create a Measurement. *Cultural Anthropology Methods 8*(3):6–8.

Dressler, William W. 1997. Culture and Patterns of Poverty. Paper delivered at the 1997 meeting of the Society for Applied Anthropology, Seattle.

Dressler, William W., Jose Ernesto dos Santos, and Mauro Campos Balierio. 1996. Studying Diversity and Sharing in Culture: An Example of Lifestyle in Brazil. *Journal of Anthropological Research 52*:331–353.

Driver, Harold E. 1938. The Reliability of Culture Element Data. Culture Elements Distributions: VIII *Anthropological Records 1*:4, Berkeley: University of California Press.

Driver, Harold E. 1962. The Contribution of A. L. Kroeber to Culture Area Theory and Practice. *International Journal of American Linguistics 28*(2), Memoir 18.

Driver, Harold E. 1965. Survey of Numerical Classification in Anthropology. In *The Use of Computers in Anthropology.* Dell Hymes, ed. Pp. 301–344. The Hague and Paris: Mouton.

Driver, Harold E. 1970. Statistical Studies of Continuous Geographical Distributions. In *Handbook on Methods in Anthropology.* Raoull Naroll and Ronald Cohen, eds. Pp. 620–639. Garden City, NY: Natural History Press.

Driver, Harold E., and A. L. Kroeber. 1932. Quantitative Expression of Cultural Relationships. *University of California Publications in American Archaeology and Ethnology 31*:211–256.

Driver, Harold E., and Karl F. Schuessler. 1957. Factor Analysis of Ethnographic Data. *American Anthropologist 59*:655–663.

Edginton, E. S. 1980. *Randomization Tests.* New York: Marcel Dekker.

Efrom, Bradley, and Robert Tibshirani. 1991. Statistical Data Analysis in the Computer Age. *Science 253*:390–395.

Ellegard, A. L. 1959. Statistical Measurement of Linguistic Relationships. *Language 35*:131–156.

Feiring, B. R. 1986. *Linear Programming. An Introduction.* Quantitative Applications in the Social Sciences, Vol. 60. Thousand Oaks, CA: Sage Publications.

Fisher, R. A. 1935. *The Design of Experiments.* London: Oliver & Boyd.

Gannotti, Mary E. 1998. The Reliability and Validity of the Pediatric Evaluation of Disability Inventory for Children Living in Puerto Rico. Ph.D. diss., Storrs: University of Connecticut.

Goodenough, Ward H. 1944. A Technique for Scale Analysis. *Educational and Psychological Measurement 4*:179–190.

Goodenough, Ward H. 1963. Some Applications of Guttman Scale Analysis to Ethnography and Culture Theory. *Southwestern Journal of Anthropology 19*:235–250.

Greenacre, M. J. 1984. *Theory and Applications of Correspondence Analysis.* New York: Academic Press.

Gujarati, D. N. 1995. *Basic Econometrics.* New York: McGraw-Hill.

Hammel, E. A. 1990. A Theory of Culture for Demography. *Population and Development Review 16*:455–485.

Handwerker, W. Penn. 1988. Sampling Variability in Microdemographic Estimation of Fertility Parameters. *Human Biology 60*:305–318.

Handwerker, W. Penn. 1989. *Women's Power and Social Revolution.* Thousand Oaks, CA: Sage Publications.

Handwerker, W. Penn. 1990. Demography. In *Medical Anthropology.* Thomas M. Johnson and Carolyn F. Sargent, eds. Pp. 319–347. New York: Praeger.

Handwerker, W. Penn. 1991. Sampling Experiments as Teaching Tools: Part II, Fishing Expeditions Usually Catch Suckers. *Cultural Anthropology Methods 3*(1):8–9.

Handwerker, W. Penn. 1993a. Gender Power Differences Between Parents and High Risk Sexual Behavior by Their Children: AIDS/STD Risk Factors Extend to a Prior Generation. *Journal of Women's Health 2*:301–316.

Handwerker, W. Penn. 1993b. Empowerment and Fertility Transition on Antigua, WI: Education, Employment, and the Moral Economy of Childbearing. *Human Organization 52*(1):41–52.

Handwerker, W. Penn. 1997. Universal Human Rights and the Problem of Unbounded Cultural Meanings. *American Anthropologist 99*:799–809.

Handwerker, W. Penn. 1998. Consensus Analysis: Sampling Frames for Valid, Generalizable Research Findings. In *Using Methods in the Field: A Practical Introduction and Casebook.* V. De Munck and E. J. Sobo, eds. Pp. 165–178. Walnut Creek, CA: AltaMira Press.

Handwerker, W. Penn, and Paul V. Crosbie. 1982. Sex and Dominance. *American Anthropologist 84*:97–104.

Handwerker, W. Penn, Julie Harris, and Jeanne Hatcherson. 1997. Sampling Guidelines for Cultural Data. *Cultural Anthropology Methods 8*:7–9.

Handwerker, W. Penn, and Danielle F. Wozniak. 1997. Sampling Strategies for the Collection of Cultural Data. *Current Anthropology 38*:869–875.

Harman, H. H. 1976. *Modern Factor Analysis*, 3rd ed. Chicago: University of Chicago Press.

Hosmer, D., and S. Lemeshow. 1989. *Applied Logistic Regression.* New York: John Wiley.

Hubert, L. and J. Schultz. 1976. Quadratic Assignment as a General Data Analysis Strategy. *British Journal of Mathematical and Statistical Psychology 29*:190–241.

Jorgensen, Joseph G., ed. 1974. *Comparative Studies by Harold E. Driver and Essays in His Honor.* New Haven: HRAF Press.

Jorgensen, Joseph G. 1995. Ethnicity, Not Culture? Obfuscating Social Science in the *Exxon Valdez* Oil Spill Case. *American Indian Culture and Research Journal 19*:1–124.

Keesing, Roger. 1994. Theories of Culture Revisited. In *Assessing Cultural Anthropology*. Robert Borofsky, ed. Pp. 301–312. New York: McGraw-Hill.

Kempton, W., J. S. Boster, and J. A. Hartley. 1995. *Environmental Values in American Culture*. Cambridge: M.I.T. Press.

Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley.

Kruskal, J. B. and M. Wish. 1978. *Multidimensional Scaling*. Quantitative Applications in the Social Sciences, Vol. 11. Beverly Hills, CA: Sage Publications.

Liem, J. H., J. G. O'Toole, and J. B. James. 1992. The Need for Power in Women Who Were Sexually Abused as Children. *Psychology of Women Quarterly 16*:467–480.

Markus, G. B. 1979. *Analyzing Panel Data*. Beverly Hills, CA: Sage Publications.

McEwan, Peter J. M. 1963. Forms and Problems of Validation in Anthropological Research. *Current Anthropology 4*:155–183.

McNabb, S. 1990. The Uses of "Inaccurate" Data. *American Anthropologist 92*:116–129.

Miller, G. A. 1956. The Magical Number Seven, Plus Or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review 63*:81–97.

Mooney, Christopher Z., and Robert D. Duval. 1993. *Bootstrapping*. Quantitative Applications in the Social Sciences, Vol. 95. Thousand Oaks, CA: Sage Publications.

Murdock, George Peter. 1971. *Anthropology's Mythology. The Huxley Memorial Lecture*. Proceedings of the Royal Anthropological Institute.

Namboodiri, Krishnan. 1984. *Matrix Algebra*. Quantitative Applications in the Social Sciences, Vol. 38. Beverly Hills, CA: Sage Publications.

Odland, John. 1988. *Spatial Autocorrelation*. Scientific Geography Series, Vol. 9. Thousand Oaks, CA: Sage Publications.

Pelto, P., and G. Pelto. 1975. Intracultural Diversity: Some Theoretical Issues. *American Ethnologist 2*:1–18.

Reynolds, Mary Stephanie. 1990. Dance Brings About Everything: Dance Power in the Ideologies of Northern Utes of the Uintah and Ouray Reservation and Predominantly Mormon Anglos of an Adjacent Uintah Basin Community. Ph.D. diss., Irvine: University of California.

Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as Consensus. *American Anthropologist 88*:313–338.

Rummel, R. J. 1970. *Applied Factor Analysis*. Evanston: Northwestern University Press.

Schroeder, Larry D., Savid L. Sjoquist, and Paula E. Stephan. 1986. *Understanding Regression Analysis*. Quantitative Applications in the Social Sciences, Vol. 57. Thousand Oaks, CA: Sage Publications.

Shweder, Richard A. 1977. Likeness and Likelihood in Everyday Thought. *Current Anthropology 18*:637–658.

Sneath, P.H.A., and R. R. Sokal. 1973. *Numerical Taxonomy*. San Francisco: W. H. Freeman.

Sokal, R. R., and P.H.A. Sneath. 1963. *Principles of Numerical Taxonomy*. San Francisco: W. H. Freeman.

Tooby, John, and Leda Cosmides. 1992. The Psychological Foundations of Culture. In *The Adapted Mind*. J. H. Barkow, L. Cosmides, and J. Toomy, eds. Pp. 19–136. New York: Oxford University Press.

Tukey, J. W. 1958. Bias and Confidence in Not Quite Large Samples. *Annals of Mathematical Statistics 29*:614.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Tylor, Edward B. 1889. On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent. *Journal of the Royal Anthropological Institution of Great Britain and Ireland 18*:245–272.

Wallace, A.F.C. 1961. *Culture and Personality*. New York: Random House.

Weller, S. C. 1987. Shared Knowledge, Intracultural Variation, and Knowledge Aggregation. *American Behavioral Scientist 31*:178–193.

Weller, S. C., and A. K. Romney. 1988. *Systematic Data Collection*. Qualitative Research Methods Series, Vol. 10. Thousand Oaks, CA: Sage Publications.

Wilkinson, Leland. 1997. *SYSTAT: The System for Statistics*. Chicago: SPSS, Inc.

Wolf, Eric R.. 1982. *Europe and the People Without History*. Berkeley: University of California Press.

Wozniak, Danielle F. 1997. Twentieth-Century Ideals: The Social Construction of U.S. Foster Motherhood. Ph.D. diss., Storrs: University of Connecticut.

Yamaguchi, Kazuo. 1991. *Event History Analysis*. Applied Social Research Methods Series, Vol. 28. Thousand Oaks, CA: Sage Publications.