Formal approaches in archaeology

GEORGE L. COWGILL

In this chapter I offer some methodological and theoretical perspectives on the current state of formal methods in archaeology. Such methods are often referred to as “quantitative,” but that term is inappropriate for several reasons. It is difficult to do anything archaeological that is not “quantitative” in the sense that counts or measurements of something are made and discussed. However, work that does not go much beyond simply tabulating and inspecting some numbers, although often very useful, is not what I have in mind. Doran (1986:21) defines formal methods as “mathematical, statistical and computer-based,” and says that “rather than being merely quantitative, formal methods are characterised by a combination of abstraction, systematisation and exactness.”

I will begin by summarizing important points in an earlier review of this topic (Cowgill 1986). The tone of my earlier chapter was predominantly critical. In this chapter I will try to get that part over with quickly, and turn, in a more positive mood, to some approaches that I think will repay more attention in the near future.

In my earlier chapter I proposed that we structure our thinking in terms of three broad categories: archaeological observations (that is, data), analytical methods (which provide connections between observations and theory), and sociocultural theory (Cowgill 1986:369). I said that mathematical and formal techniques mostly belong in the category of analytical methods, and while there is room for improvement in these techniques, what is most needed is better sociocultural theory, improved data, and better understandings of the connections between archaeological evidence and past behavior (that is, better “middle-range” theory) (Cowgill 1986:370, 390).

One set of commentators interpreted this as only a call for better theory and ignored my emphasis on the need for better data as well (Meltzer et al. 1986:14). This suggests that many archaeologists have a mind-set that makes them complacent about the quality of their data, so that they focus their attention on new theory and have trouble “hearing” the message that the quality of the data themselves is often problematic.

To be sure, at least one reviewer understood my point about data (Fagan 1986:211–12), and perhaps I exaggerate the problem. Nevertheless, I will repeat it in a way that should be impossible to ignore. It is a variation on the old themes “garbage in, garbage out” and “You can’t make a silk purse from a sow’s ear.” These statements do not mean that if we don’t have nearly perfect data we should simply give up. The message is that archaeological data are usually marginal in quality,
relevance, and quantity for most of the really interesting topics in sociocultural theory. We are unavoidably in the business of refining low-grade data to win something of value. One can think of it as tuning our detection gear ever so finely to try to decode voices that are faint, alien, distorted, and almost drowned out by irrelevant messages and random noise. To do this very well we need all the help we can get. Formal approaches are one very important source of help, but they do not substitute for anything else. My point about data was, and remains, that more or less elaborate analyses (non-formal as well as formal) are often applied to data bases whose weakness does not seem to be a matter of concern. In an era of rapid site destruction and scarce funding, such data are sometimes the best we can hope for. But insensitivity to problems about quality of data will do nothing toward improving the situation, and awareness of the problem is at least a beginning.

In my 1986 paper I noted several specific kinds of trouble with data that can be alleviated by better methods. It is a mistake to think that a survey is ever “total”; more intensive survey will always identify additional archaeological occurrences. I argued further that even the best spatial sampling schemes are inferior substitutes for doing a survey that is as intensive and comprehensive as possible, especially because spatial sampling is apt to miss occurrences that are inconspicuous but very important and because spatial sampling makes it almost impossible to correctly recognize spatial patterns unless they are much smaller than the sampling units employed. I do not think spatial sampling can be regarded as a way to avoid unprofitable redundancy in regional survey, because it is highly probable that fuller survey will yield unexpected important new information. Sampling should be used for regional survey only when one is forced to it by shortage of time and resources. Nance (1983) provides a very useful review of regional sampling.

I also pointed out that it is a mistake to think that a collection is ever “total.” There must always be some cutoff below which objects are ignored because they are considered too unimportant (usually because they are too small and featureless). The composition of collections can be significantly affected by these cutoffs. Since we cannot really collect everything, we should use consistent criteria for what to ignore, and report these criteria routinely and carefully (Cowgill 1986:382–3).

I criticized the term “controlled collection” because it is often used to suggest that collection procedures were highly reliable – that is, that what one field worker collected one day from a site is highly similar to what would have been collected on any other day by any other member of the same project. In fact, the term is nearly meaningless unless one carefully explains what steps were taken to achieve high reliability (Cowgill 1986:382–3). These steps should include some repeated collections made by different workers under different circumstances from the same sites, in order to provide an empirical basis for assessing reliability.

I was especially critical of the notion that “around 100 objects” is a good target for collection sizes. Unless the material in question is remarkably lacking in diversity, some categories are bound to be represented by small numbers in such collections. For example, if there are twenty categories of interest, all approximately equally common, then on the average each will be represented by about one-twentieth of each
collection, or about 5 objects in a collection of 100 objects. One is then faced with the
difficult problem of deciding whether a count of 2 or 8 really means something, since
deviations that large can easily occur by chance when the true proportion (in the
population of objects sampled by a specific collection) is really $\frac{1}{2}$.
If instead one had a
collection of 400 objects, obtained as an approximately random sample from a
population where the true proportion of a given category is $\frac{1}{2}$, one would expect a
count of around 20 of that category in the collection. An observed count of 8 would
provide strong evidence that the true population proportion was distinctly less than
$\frac{1}{2}$ and was more likely around $\frac{1}{50}$, while an observed count of 32 would provide
strong evidence that the proportion was greater than $\frac{1}{2}$ and more likely around $\frac{8}{100}$.

A further problem with a small collection is that it is likely to underrepresent the
ture diversity of the population represented by the collection (Jones et al. 1983;
Kintigh 1984; Jones and Leonard in press). Comparisons of assemblage diversity are
misleading unless collection size is taken into account, and larger collections give
more reliable estimates of true diversity. Finally, correlations between either counts
or percents of categories are not merely unreliable, they are systematically biased
toward zero if they are based on data from collections so small that there are never
more than a few objects belonging to the relevant categories in any one collection. For
all these reasons, collections of only around 100 objects are usually hard to deal with
statistically. Often we have no choice and must make the best use we can of small
collections. But it would be better to make collections of several hundred objects, if
this is at all feasible (Cowgill 1986:383-4).

I ended my criticism of data quality by saying that excavation techniques still often
leave much to be desired. Few will admit to ignoring distinct layers when they are
observed, but there are still tremendous differences in the extent to which care and
skill in recognizing layers is emphasized in archaeological training and practice. There
are also sometimes surprising deficiencies in drawing profiles and other aspects of data
recording. Elaboration of formal methods (or of archaeological theory) does not mean
that we are somehow beyond all that tedious attention to excavation technique. It
means, instead, that there is more need than ever for good technique, and more point
to it (Cowgill 1986:385-6).

In an earlier section of the paper I talked about the relationship between archae­
ological data and sociocultural theory. Although there are a number of formal
techniques that are intended to aid in building models that parsimoniously and
effectively describe data, it is incorrect to equate this process with the building of
explanatory theory. It is, instead, only one component, albeit a very important one, in
the creation, testing, and improvement of theory (Cowgill 1986:371-8).

At various places in the paper I noted that formal methods have been used for the
design of research (data collection), for summarization and exploration of data, and as
aids in efforts to connect data with sociocultural theory, but that there was still
remarkably little relevant sociocultural theory that was itself expressed in formal
terms. Many of the most important concepts in theory seem fully expressible by very
simple formal relationships (such as “greater than” and “less than,” “highly probable”
and “quite improbable”). What is most important are concepts coming from sources
such as ethnography, social anthropology, and social history. I said that “richness of mathematical concepts cannot remedy poverty of sociocultural concepts” (Cowgill 1986:374). This is worth emphasizing because, after having just argued that we must be superb technicians, I am now arguing that we must also be widely and perceptively acquainted with knowledge that is only remotely connected with the materials we study directly. If this seems too much for any one person, I can only reply that we must do it, or at least people who do different parts of it must communicate in genuinely effective ways, if archaeology is to progress.

Attempts to express sociocultural theory in formal terms are commonly vitiated by the shallowness, naivety, or ethnocentrism of the sociocultural “data” and assumptions embodied in the formal model. This deficiency is by no means distinctive of archaeologists. Many of the worst offenses are committed by economists and political scientists, who often show a quite bizarre willingness to base their understandings on official statistics and interviews with elites, even when they could collect the kind of ethnographic data archaeologists would give their eye teeth for. Rather than “thick” description, they seem content with descriptions so thin that they border on the ectoplasmic.

Is the “thinness” and ethnocentrism of the inputs characteristic of formal models an unavoidable consequence of the abstraction these models require, or is it something else? The ethnocentrism could be considerably reduced by proper regard for ethnographic knowledge. Would this suffice for the production of formal models that are abstract and yet apt and relevant enough to be of value? Or is the problem, which is that of taking adequate account of ideational realms that are rich and powerful but very different from the ideational realm within which the formal models are created, very much more intractable? Many readers may already have their minds made up on this issue. But to me it is an open question, and one that I pose as a challenge.

With this, I turn to several specific topics to which I want to call attention because, in my judgment, it will be profitable in the near future to pursue work along these lines. I emphasize especially matters omitted from my earlier chapter. Several of these are still relatively unknown to archaeologists.

**Bayesian approaches**

The “Bayesian” approach is named after Thomas Bayes, author of a theorem basic for its formalization that was published posthumously in 1763. The mood of the approach is, however, perhaps best introduced by a relatively informal example. The date of the earliest human occupation of the New World is controversial. Some archaeologists doubt if it was much more than 12,000 years before the present, while others favor a far earlier date. On brief reflection and introspection, I find that I can quite comfortably sketch my own current opinions (though certainly not the reasons for those opinions) by means of a simple table.

The second column of Table 5.1 can be thought of as a basis for the betting odds I would give at present for different possible dates, if I were a betting person. Other archaeologists who have given any attention at all to the problem are not likely to agree entirely with my probabilities. Some may even insist that they cannot meaning-
Table 5.1 Personal prior probabilities of the author and two hypothetical other archaeologists ("E" and "L") for the presence of humans in the New World by various dates

<table>
<thead>
<tr>
<th>Date (years B.P.)</th>
<th>GLC</th>
<th>E</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000</td>
<td>0.01</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td>20,000</td>
<td>0.05</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td>15,000</td>
<td>0.90</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>12,000</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Note: Probabilities are rounded to two decimal places. Probabilities shown as 0.00 and 1.00 do not necessarily mean absolute certainty.

fully assign any numbers at all to their own views. However, I think it unlikely that even they will remain neutral about Table 5.1. Some will feel quite strongly that I have assigned much too low a probability to the earlier dates. Others may feel that my probabilities are about right. At least a few will think I have given too high a probability to early dates. We can imagine two other archaeologists; E, who favors an early date, and L, who is even more skeptical of an early date than I am. The probabilities that reflect the views of E and L are summarized in columns 3 and 4 of Table 5.1.

Suppose a new site is found in North America, with really good evidence of human occupation and several radiocarbon dates that are consistent, widely agreed to be unproblematically associated with the human occupation, and cluster around 20,000 B.P. Even one such case without any “weak link” in the evidence would suffice to cause me to revise considerably the views represented in Table 5.1, which is a prior probability distribution (“prior” because it is prior to taking into account some specific body of new evidence). My posterior probability for humans in the New World by 20,000 B.P. might jump from 0.05 to as high as 0.95. There are many instances where skeptics have seen some body of really strong evidence and changed their minds. This happened, for example, in the late 1920s when American archaeologists first became persuaded of the reality of the human occupations that are now dated around 10,000 to 12,000 B.P.

Other archaeologists will react differently to the new data. E was already practically certain people were here 20,000 years ago, and his/her opinion will be unaffected. L, however, might still be unwilling to assign a higher posterior probability than 0.05 for humans in the New World by 20,000 B.P., and might think it more likely that “something is wrong” with the data on the newly reported site.

All this may seem woefully unscientific. Surely we are all supposed to be “objective” and to draw identical conclusions from any particular body of evidence. If we allow the opinions we held prior to examining the evidence to influence our subsequent opinions, do we not open the door to anarchy?
The answer many statisticians have given to this question is that to pretend that our prior views do not influence what we think the data mean does not enable us to be objective; instead it confuses us about what we really do, and encourages us to reason illogically and ineffectively. Furthermore, if our prior views are not wholly incompatible, examination of a given body of new evidence will lead to posterior views that are less divergent than the prior views were.

This last proviso is important, and I think it could be paraphrased by saying that for people’s posterior views to converge, they must begin by at least sharing the same paradigm. For example, it is not likely that any amount of new evidence will lead to convergence between the views of “creation scientists” and evolutionary biologists, because they begin with incompatible assumptions. On the other hand, archaeologists who argue heatedly about whether humans were in the New World long before 30,000 B.P. or not before 12,000 B.P. conduct their arguments in the context of a large body of shared assumptions. When there is broad agreement about the nature of reality and the modes of reasoning that are valid, then it is possible for disagreements on specific points to be resolved, or at least greatly reduced, by consideration of new evidence.

It may be objected that people don’t really behave like the hypothetical archaeologist L. In particular, L may actually be far less willing to change his or her mind than the prior probability of 0.01, which is small but not extremely small, would suggest. But if that is the case, then either L was not realistic about his or her prior probability, or there is a clear sense in which L is behaving illogically. As expressed in qualitative terms, the “Bayesian mood” seems to me partly a way of thinking about how we really behave, and partly a sensible prescription for how we ought to behave, at least for “normal science.” Kuhnian revolutions are perhaps another matter.

One possible misunderstanding needs to be cleared up at once. The viewpoint that I advocate is totally opposed to the notion that there is no such thing as absolute truth, and that any person’s opinion is just as good and just as “true” as anyone else’s. To be sure, there are any number of different socially constituted realities, and we fall short in understanding ourselves to the extent that we cannot enter the ideational realms of others and understand their own underlying assumptions, logic, and preoccupations. Nevertheless, a deep assumption of my native culture is that there is something “out there” that exists and has properties independently of our awareness of it. How we perceive what is “out there” is the result of a complex interaction between what is outside us and the very complex apparatus of prior knowledge and sensitivities by which we experience things. I believe it is possible to construct models of the world that increasingly approximate how it really is, even if we can never get beyond approximations. I also think that doing science is a distinctive and unusually effective way of improving our approximations, in large part because doing science involves systematically subjecting our ideas to challenge and being prepared to change ideas that do not stand up to these challenges. The “Bayesian mood” sketches some good guidelines for ways to modify ideas. It is, thus, not a denial that there is something definite “out there,” but rather a way to facilitate the approximation of our models to that something definite.
To make this argument more specific, if we consider any region of the New World, I have no doubt that there was a definite date when the first humans set foot in it. It is highly unlikely that we will ever be able to determine such a date with complete certainty and accuracy. Nevertheless, it is both likely and desirable that archaeologists will arrive at a far narrower range of estimates agreed to be plausible than is the case at present.

In this first example, it does not seem to be feasible to replace what I have called the "Bayesian mood" with specific Bayesian techniques, because, even if we had the hypothetical new site with "very good" evidence of occupation at 20,000 B.P., it is not clear how that "very good" evidence could be equated with some specific numerical probability. In other cases, however, there are reasonable ways to do this. I will illustrate this with a second example. In the nineteenth century the "Long Count" of the Classic Maya calendar was deciphered, and this made it possible to determine with great confidence the exact number of days between two dates recorded by the Maya. However, the correlation between the Long Count and the European calendar remained controversial for many years and is still not settled to the satisfaction of all competent scholars. For technical reasons only a limited number of correlations are plausible, and those considered most plausible occur at intervals of about 256 years. For simplicity, consider just the two most popular, which I will call "early" and "late." Suppose a stela is found, with a Long Count date that correlates either with A.D. 400 (early) or A.D. 656 (late). Suppose also that a calibrated radiocarbon date of A.D. 600 is obtained from material that indisputably ceased to metabolize at just about the time recorded by the stela date. Unfortunately, the date was determined by antiquated equipment and it comes with a standard error of 250 years. For purposes of illustration, assume that the probability distribution of the date nearly follows a Normal (i.e. Gaussian) curve (the distribution is actually Poisson, which is slightly different). A standard statistical approach would then be to say that the late correlation implies a date that is (656 - 600)/250 or 0.224 standard errors later than the "best" radiocarbon estimate. Consultation of the table of areas under the Normal curve that is to be found in every introductory statistics text shows that a deviation as large as this, or larger, is to be expected in about 82 percent of cases where the hypothesis (that the true date is A.D. 656) is correct. On the other hand, the "early" correlation implies a date that is (600 - 400)/250 or 0.80 standard errors earlier than the radiocarbon estimate. The probability of a deviation this large or larger is about 42 percent. Another approach is to compute a "confidence interval" about the radiocarbon date. For example, statistical theory tells us that if we construct an interval that extends 1.645 times the standard error above and below the best estimate, 90 percent of such intervals will include the true value. In this case, 1.645 x 250 = 411, and we get a 90 percent confidence interval that runs from A.D. 189 (600 - 411) to A.D. 1011 (600 + 411). This interval, of course, comfortably includes both A.D. 400 and A.D. 656.

By any correct way of looking at it, the radiocarbon date has not helped noticeably. To be sure, 600 is considerably closer to 656 than to 400, but the very large uncertainty in the date means that accepting the radiocarbon evidence does not
require one to regard either correlation as particularly implausible. Perhaps the greatest danger is that a statistically naive advocate of the late correlation might argue that since the radiocarbon date is highly consistent with the late correlation, it adds meaningful support to that correlation. But a naive advocate of the early correlation could say that the date is highly consistent with that correlation, and therefore highly supportive of it. The fact is, of course, that the date comes with such a large standard error that it is highly ambiguous, and thus provides almost no evidence one way or the other. For such poor evidence, it may be reasonable to leave it at that.

Suppose now that the radiocarbon date is redetermined with better equipment and that it happens to yield the same value, A.D. 600, but now with a standard error of only 80 years. The late correlation date is (656 - 600)/80 = 0.70 standard deviations high, and if A.D. 656 were the true date a deviation this large or larger is to be expected in about 48 percent of samples. However, the early date is (600 - 400)/80 = 2.50 standard deviations low, and if A.D. 400 were the true date, a deviation this large or larger would be expected in barely more than 1 percent of samples. The 90 percent confidence interval spans the range A.D. 468 to 732, and even the 95 percent interval is only from A.D. 443 to 757. This new evidence considerably favors the late correlation. According to the prevailing custom of treating the 5 percent significance level as an unerring guide for decisions, we could say that the hypothesis that the early correlation is correct can be rejected at the 5 percent level, and there is an end to the matter . . . or at least an end until new evidence appears.

But is this really the end? Should advocates of the early correlation give up that easily? Will they, whether they should or shouldn’t? According to a common but incorrect understanding of the “hypothesis testing” approach, the late correlation must be “provisionally accepted” and the early correlation rejected. More correctly, advocates of the early correlation are put on the defensive but not required to abandon their position. They ought perhaps to hold it with less confidence than before, but conventional statistical analysis gives no guidance beyond that vague suggestion (Cowgill 1977).

A Bayesian approach offers considerably more help. If one can give a rough quantification of one’s prior probabilities for the competing hypotheses, then the radiocarbon evidence can be used to give an unambiguous set of posterior probabilities. This was not possible in the earlier example because the new evidence could not be expressed in a definite quantitative form. In the present example, the mean of the radiocarbon date, its standard deviation, and the fact that the distribution is approximately Normal are sufficient information to permit the computation of definite posterior probabilities. To illustrate this, imagine again two archaeologists, A and B, whose prior probabilities are shown in column 2 of Table 5.2. A’s prior probabilities are 0.10 for the early correlation, and 0.90 for the late. That is, s/he leans toward the late, but figures there is about one chance in ten that the early correlation may be correct. B, on the other hand, holds prior probabilities of 0.99 for the early correlation and only 0.01 for the late correlation. B is a very strong advocate of the early correlation and quite skeptical of the late correlation, though willing to admit that it just might be correct. By simple computations that take only a few minutes by
Table 5.2 Prior and posterior personal probabilities for two individuals ("A" and "B") concerning two possible correlations of the Maya and European calendars, before and after taking into account the evidence of a single new radiocarbon date

<table>
<thead>
<tr>
<th>Individual</th>
<th>Prior probability for &quot;late&quot;</th>
<th>Posterior probability after taking into account a radiocarbon date of A.D. 600 with standard error of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>250 years</td>
</tr>
<tr>
<td>A</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>B</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: Probabilities for the "late" correlation are shown. Probabilities for the "early" correlation can be found by subtracting "late" probabilities from one.

pocket calculator, the result of taking into account the first radiocarbon determination, with a standard error of 250 years, leads to posterior probabilities for A of about 0.08 for the early correlation and 0.92 for the late correlation. For B, the posterior probabilities come out to about 0.99 and 0.01. In other words, the radiocarbon evidence was so ambiguous that it has had almost no impact on the prior probabilities. However, the effect of a radiocarbon determination with a standard error of only 80 years is very different. It leads to posterior probabilities for archaeologist A of 0.01 and 0.99. This is a substantial further shift in favor of the late correlation. The posterior probabilities for B are 0.85 for the early correlation and 0.15 for the late. This is also a substantial shift toward the late correlation, but B could rationally still give odds of 0.85 to 0.15, that is, 17 to 3, in favor of the early correlation.

At this point it may seem that the Bayesian approach is little better than the conventional approach, since A and B began by disagreeing and they still disagree. But suppose the radiocarbon date had a standard error of only 50 years. Bayesian computations show that A can now feel virtually certain that the late correlation is correct, while the posterior probabilities for B are about 0.06 for the early correlation and 0.94 for the late correlation, as shown in the last column of Table 5.2. In other words, the rational thing for a person who holds B’s prior probabilities to do, in the light of the radiocarbon evidence, is to switch from a very strong belief in the early correlation to moderately strong belief in the late correlation.

All these posterior probabilities can be used as prior probabilities whenever further evidence is obtained, in a cyclic interaction between beliefs and further evidence. The stronger the evidence, the greater the narrowing of the gap between initially diverse personal views.

A real archaeologist who originally assigned a probability of 0.01 to the late correlation, when confronted with a single radiocarbon date of A.D. 600 with a standard error of 50 years, might not actually be willing to switch to a posterior probability of 0.94 for the late correlation. However, this unwillingness could only be based on (a) hitherto unexpressed suspicions that the true uncertainty in the radiocarbon date was much greater than the 50-year standard error implies, (b) an
admission that his or her true prior probability was really considerably less than 0.01, or (c) indefensible irrationality.

This is only a sketch of one way in which Bayesian methods can be used, to avoid the obfuscation of our actual reasoning processes encouraged by the "hypothesis testing" approach, and to help us make our reasoning more explicit, more coherent, and more powerful.

There are many other possible applications and many specific Bayesian techniques. "Empirical" Bayesian methods can be used, for example, to make estimates of population parameters that are better than those made by more conventional approaches. I believe that Chernoff (1982) is the first to have applied empirical Bayesian techniques to an archaeological problem. Iversen (1984) gives a brief and relatively simple introduction to Bayesian statistical inference. Box and Tiao (1973) is a more advanced text on Bayesian methods.

Activity signatures and intrasite spatial analysis
Figuring out the activities represented by an archaeological occurrence is basic for good interpretation. If the occurrence covers more than a tiny area, it is also desirable to infer as much as possible about the spatial patterning of activities. Ideally, there should be good diagnostics of different activities. By a "good diagnostic" I mean a kind of evidence that is relatively durable and is consistently left behind by a certain type of activity in quantities sufficient so that more than a very little bit of the activity will practically always leave behind a detectable amount of evidence. It is also important that no other type of activity should generate much, if any, of this kind of evidence. Examples of things that are sometimes good diagnostics are specific artifact categories, distinctive kinds of waste or by-products, discarded tools with distinctive types of wear or damage, special kinds of immobile facilities or features, and distinctive alterations to land surfaces or layers (such as high-temperature burning). If there is at least one good diagnostic for an activity, then the activity leaves behind a relatively clear "signature." If the diagnostics are immovable or if there is reason to think they have not been moved from the spot where the activity took place, they also provide fairly straightforward evidence about the spatial pattern of the activity. If there is reason to think that the good diagnostics were tossed, dumped, or otherwise moved away from the spot where the activity took place, it is harder to tell much about the spatial pattern of the activity but it can at least be inferred that the activity took place nearby. Specialized tools, even if they are used for only a single type of activity, cannot be good diagnostics if they are usually "curated" and rarely discarded or lost at the spot where the activity took place.

In the absence of identified good diagnostics, can anything useful be done? Much effort has been devoted to aspects of this problem. Noteworthy recent examples include Kintigh and Ammerman (1982), Whallon (1984), and many of the chapters in Carr (1985). I think the model that is at least implicit in most of this work is that activities are distinguishable by different proportions of different types of remains. That is, although no one category of remains is a good diagnostic of any of the activities, at least some activities are distinguishable from other activities by propen-
Table 5.3 Example of a situation where there are no “good diagnostics” of any activity, but different propensities to leave behind different proportions of different categories of remains leave contrasting signatures of three activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Category v</th>
<th>Category w</th>
<th>Category x</th>
<th>Category y</th>
<th>Category z</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>b</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>c</td>
<td>15</td>
<td>5</td>
<td>40</td>
<td>10</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 5.3: Example of a situation where there are no “good diagnostics” of any activity, but different propensities to leave behind different proportions of different categories of remains leave contrasting signatures of three activities.
another in the mixes of activities represented, the indeterminacy in the solution could be made small enough so that the result would be quite useful.

This was as far as I was able to carry the matter when (thanks to Ron Bishop of the Smithsonian Institution) I discovered that there is already a sizable literature by geologists and mathematicians concerned with an analogous problem, referred to as “linear unmixing.” This deals with the situation where a data set “can be viewed as a collection of samples representing mixtures of a relatively small number of end members” but where the end members themselves are not necessarily present in the set of samples (Full et al. 1982). What they call “end members” are the same as what I call “activity signatures” here. To say that an end member is not present in the sample is the same as to say that no site segment represents the pure signature of a given activity. Full et al. (1982) use a “fuzzy subsets” technique (Zadeh 1965) to deal with the problem of noisy data – essentially to make the reconstruction of end members considerably more robust in the face of data errors, samples that do not belong (e.g. a site segment that actually pertains to a culturally different occupation), and so on. So far as I know, linear unmixing has not yet been applied to archaeological data. However, I think it by far the most promising solution available to the problem of inferring activity signatures in the absence of good diagnostics.

This is not to say that the linear unmixing approach is guaranteed to give good results. For one thing, the whole concept of activity signatures remains shaky. If there is not much consistency from one time to the next in the proportions of things left behind by people engaged in a certain type of activity – in other words, if the variances in proportions actually left behind in specific instances are too large relative to differences between average proportions left behind – then linear unmixing probably will not give good results. However, if activities do not leave reasonably consistent signatures, I cannot imagine any other approach that would work better.

Linear unmixing should be able to deal with situations where a fraction of the material left behind by an activity is left in place and another fraction is tossed or dumped. All that is needed is that the propensities for dropping, tossing, and/or dumping should not have variances that are too large. If that condition is satisfied, one could identify a “dropping signature” for a given activity, a “dumping signature” for the same activity, and so on.

The dilemma of whether to use counts or proportions afflicts R-mode principal components approaches. Using raw counts causes correlations to be undesirably influenced by differences in absolute quantities of material in different site segments, while standardizing by converting to proportions introduces complex constraints in correlations because the proportions in a segment must sum to unity. A very attractive feature of linear unmixing is that it is untroubled by this problem.

Finally, linear unmixing is probably fairly insensitive to the size and shape of the site segments used. The signatures of different activities in a site may result in scatters of material that are very diverse in size, shape, and density (Whallon 1984) and this has been a vexing problem for efforts to infer activities and their spatial distribution, if no good diagnostics are identified. For linear unmixing, it is only necessary that the site segments represent a fairly wide range of proportionate mixes of the signatures of
different activities. It is not necessary that a site segment be homogeneous in the
signatures present, or that the boundaries of the segment have any relationship to the
edges of any particular signature. If site segments are too large they will not differ
enough from one another for linear unmixing to work well, and also the spatial
resolution will be poor. If segments are too small, they will tend to contain very small
quantities of material and random statistical errors will be too large. But it seems that,
between these extremes, there will be a fairly wide range of sizes and shapes of site
segments for which linear unmixing will work well, if well-defined activity signatures
exist at all. It would be logical to use segments as small as they can be made without
making random errors in counts within each segment too large. I see no reason why all
the segments have to be the same size. If linear unmixing enables one to infer activity
signatures reasonably well, the proportion of a given signature in each site segment
could be plotted (like the factor scores in a principal components study) in order to
show the spatial distribution of that activity signature. The size and shape of the
signature would emerge naturally from such a plot, with a resolution determined by
the sizes of the site segments.

Typology and systematics
Most formal approaches to typology or classification have been preoccupied with the
problem of finding a way to divide a set of objects into subsets that are in some sense
optimal clusters or groups. The general idea is that each subset should consist of
objects that are relatively similar to one another and relatively different from all the
objects in other subsets. This problem is interesting mathematically and there is a
large literature on it, much of it in biology and other non-archaeological fields, which
attests to the belief that it also has great practical importance. Undoubtedly this is so,
and continued archaeological pursuit of this problem seems worthwhile. But I urge
that we also work on the formalization of other aspects of systematics. I am thinking
of situations where objects differ enough so that they unquestionably belong in
different classes (in the sense of Dunnell 1971, 1986), yet show a "family resem­
blance" that cannot plausibly be accidental. For example, cylindrical vases with flat
bottoms, vertical or slightly concave walls, and slab-like tripod supports are very
widespread in Mesoamerica between about A.D. 200 and 600. They share a general
resemblance that is surely not accidental, and their presence at a site has sometimes
been interpreted as reflecting direct and strong "influence" from some single center,
such as Teotihuacan. More than superficial study of the objects shows, however, that
there are a number of regional and temporal variants, as well as great differences in
pastes and techniques and styles of decoration. It is likely that closer comparison of
many examples would reveal complexly intertwined patterns of imitation and inno­
vation, with interesting implications about interactions between Mesoamerican
societies. However, attempts to optimize discovery of the "best" clusters do not seem
highly relevant for unravelling these relationships. To be sure, clusters can be
organized hierarchically into clusters of clusters, clusters of clusters of clusters, and so
on, but these hierarchies themselves are rigid and can only express certain resem­
blances at the expense of others. The "fuzzy subset" concept (Zadeh 1965; Bezdek
1981) may be helpful, since it does not require an object to be wholly a member of only one cluster. Instead, it can be, for example, 70 percent a member of cluster A, 20 percent a member of cluster B, and 10 percent a member of cluster C. However, this still requires that the intricate pattern of partial resemblances between an object and a subset of other objects be reduced to a single number, and this is probably too drastic an abstraction from the observed data.

In order to deal adequately with these "family resemblance" problems it is clearly necessary to do a good deal of painstaking and rather old-fashioned comparison, relying far more on specific attributes and "modes" than on class definitions. It may be that that will be sufficient, as well as necessary. I suspect, however, that there are ways to formalize such investigations that will be very helpful. For example, the "design grammar" approach (Chippindale and Hassan, in preparation) offers ways to formalize the description of local styles, and comparisons of such styles will probably be facilitated by phrasing them as comparisons between the design grammars.

The problem of identifying patterns of imitation and innovation within a corpus of material that is obviously related (though perhaps often distantly) seems to have much in common with that of working out the relationships between different manuscript versions of an ancient text, and even more in common with the interplay of diverse influences and new ideas in the history of an artistic style. All this points vaguely in the direction of data bases with sophisticated and flexible organizations. Some steps in this direction have already been taken (e.g. Gardin 1980; Langley 1986). I urge that much more attention be given to such approaches.

**Some brief notices**

Space does not suffice to deal adequately with all important formal approaches in archaeology. There are three more, however, that require at least brief mention.

**Discrete multivariate analysis**

Cross-tabulations of pairs of nominal scale variables and associated statistics, such as chi-square, are relatively familiar. Extensions of this approach to consider the joint effects and interactions of a number of nominal scale variables has flourished only in the past two or three decades. In spite of a number of archaeological papers on the methods (Spaulding 1976, 1977; Read 1974; Clark 1976; Lewis 1986), there have as yet been relatively few archaeological applications. This is perhaps because even with only four or five variables the simplest models that satisfactorily summarize the data may seem disturbingly complex, and it may also be that many archaeologists have not seen, from the examples presented, how to make the approach bear directly on the problems that most interest them. I predict, however, that in the near future discrete multivariate analysis will find many applications in archaeology.

**Exploratory data analysis**

This label refers to a diversity of techniques that share the attitude of looking without too many preconceptions at a "batch" of data in various ways to try to see what structure or pattern may lurk within, and a concern with "robust" methods that are
not much disturbed by a few aberrant or erroneous data values or by serious departure of the data from traditionally popular assumptions such as approximate Normality. Many of the techniques offer very welcome alternatives to "standard" procedures described in elementary statistics texts, and with good reason they are becoming very popular in archaeology. They are increasingly available in common statistical systems for computers. They do not, of course, supersede other techniques that require stronger assumptions about the data and "cleaner" data, but they are extremely valuable additions to the repertoire of techniques, especially in earlier stages of an analysis. Many of the techniques and concepts are relatively simple, and all archaeologists should become acquainted with them. Tukey (1977) and Mosteller and Tukey (1977) are good introductions.

Artificial intelligence and expert systems
I have had no direct experience with this field, and I rely largely on a recent discussion by Doran, who argues that these non-numerical but formal methods will be particularly effective in suggesting "new forms of theory, both sociocultural and middle-range, precisely because they already, and in practice, bridge the gap between data and process models" (Doran 1986:32). Clearly a great deal more work needs to be done, and Doran cautions that "progress will not be easy" (Doran 1986:32). But, to return to the Bayesian mood, I think the odds that something important will come from this direction are good enough to warrant careful attention to it.

Notes
1 Binford (1986:463-4) charges that some archaeologists seem to think of themselves as merely disadvantaged ethnographers. Clearly we should not think of ourselves that way. To do so carries the tacit implication that if only we had access to the kinds of information on ancient societies that an ethnographer can get on contemporary societies, all our questions would be answered, whereas the magnitude of unsolved theoretical questions about ethnographically accessible societies is only too apparent. Furthermore, archaeologists can get kinds of information, especially about changes over time, that are rarely available to ethnographers. Nevertheless, the kinds of highly relevant information that we cannot get, or can only infer with great difficulty, remain very frustrating. Our understanding is greatly aided when archaeological data on a society can be combined with documentary and/or ethnographic data on the same society.

2 This use of the term "family resemblance" comes from Wittgenstein and has been taken up in social anthropology by Needham and others (Benson Saler, personal communication).